

Mitogenomes of Accipitriformes and Cathartiformes Were Subjected to Ancestral and Recent Duplications Followed by Gradual Degeneration

Adam Dawid Urantówka^{1,*}, Aleksandra Krocak^{1,2}, Tomasz Strzała¹, Grzegorz Zaniewicz³, Marcin Kurkowski¹, and Paweł Mackiewicz ^{2,*}

¹Department of Genetics, Wrocław University of Environmental and Life Sciences, Poland

²Department of Bioinformatics and Genomics, Faculty of Biotechnology, Wrocław University, Poland

³Department of Vertebrate Ecology and Zoology, Avian Ecophysiology Unit, University of Gdańsk, Poland

*Corresponding authors: E-mails: adam.urantowka@upwr.edu.pl; pamac@smorfland.uni.wroc.pl.

Accepted: 15 August 2021

Abstract

The rearrangement of 37 genes with one control region, firstly identified in *Gallus gallus* mitogenome, is believed to be ancestral for all Aves. However, mitogenomic sequences obtained in recent years revealed that many avian mitogenomes contain duplicated regions that were omitted in previous genomic versions. Their evolution and mechanism of duplication are still poorly understood. The order of Accipitriformes is especially interesting in this context because its representatives contain a duplicated control region in various stages of degeneration. Therefore, we applied an appropriate PCR strategy to look for duplications within the mitogenomes of the early diverged species *Sagittarius serpentarius* and Cathartiformes, which is a sister order to Accipitriformes. The analyses revealed the same duplicated gene order in all examined taxa and the common ancestor of these groups. The duplicated regions were subjected to gradual degeneration and homogenization during concerted evolution. The latter process occurred recently in the species of Cathartiformes as well as in the early diverged lineages of Accipitriformes, that is, *Sagittarius serpentarius* and *Pandion haliaetus*. However, in other lineages, that is, *Pernis ptilorhynchus*, as well as representatives of Aegypiinae, Aquilinae, and five related subfamilies of Accipitriformes (Accipitrinae, Circinae, Buteoninae, Haliaeetinae, and Milvinae), the duplications were evolving independently for at least 14–47 Myr. Different portions of control regions in Cathartiformes showed conflicting phylogenetic signals indicating that some sections of these regions were homogenized at a frequency higher than the rate of speciation, whereas others have still evolved separately.

Key words: Aves, duplication, ancestral state, concerted evolution, control region, gene order, mitochondrial genome, mitogenome, Accipitriformes, Cathartiformes, rearrangement, phylogeny.

Significance

Mitochondrial genomes of vertebrates have been considered stable in its structure for a long time. However, thanks to a new PCR strategy, we identified previously omitted duplicated regions in the mitogenomes from related avian orders, Accipitriformes (the diurnal birds of prey) and Cathartiformes (the New World vultures). Our study revealed that their control regions were subjected to dynamic processes: duplication, homogenization via concerted evolution, and gradual degeneration. Moreover, we found that different portions of control regions in Cathartiformes show conflicting phylogenetic signals indicating that some sections of these regions were homogenized at a frequency higher than the rate of speciation, whereas others have still evolved separately. These findings shed new light on the mechanism of evolution of the mitogenomic duplications.

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Introduction

Most of avian diurnal raptors are grouped into the order Accipitriformes comprising 249 extant species classified into 70 genera (Clements et al. 2019). This group is sister to New World Vultures, that is, Cathartiformes (Jarvis et al. 2014; Prum et al. 2015; Clements et al. 2019; Knapp et al. 2019). Both of them were lumped into the group of Accipitrimorphae (Jarvis et al. 2014), which occurred related to Strigiformes in recent analyses (Houde et al. 2019). Nearly all Accipitriformes taxa are placed in the family Accipitridae, whereas Sagittariidae and Pandionidae families are represented only by *Sagittarius serpentarius* and *Pandion haliaetus*, respectively. Interestingly, the taxonomic division into the three families is reflected in the three different mitochondrial gene orders that have been identified in representatives of these families.

Sagittarius serpentarius, which is the earliest diverged species of Accipitriformes (Mindell et al. 2018; Knapp et al. 2019), contains a mitogenome (Mahmood et al. 2014) with the gene order identical to *Gallus gallus* (Desjardins and Morais 1990) (fig. 1). Since the chicken-type gene order prevailed in known avian mitogenomes, it was assumed to be ancestral for all Aves and called typical, standard, or common. This rearrangement derived likely from the typical vertebrate gene order by a single tandem duplication of a fragment located between *ND5* and *tRNA-Phe* genes followed by random losses of one copy of duplicated genes and control regions (Schirtzinger et al. 2012). So far, different rearrangements including the duplicated control region and genes were identified in representatives of most avian orders (Eberhard and Wright 2016; Urantowska et al. 2018, 2020; Mackiewicz et al. 2019). The most fully duplicated gene order, assigned as GO-FD in figure 1, was found in the mitogenomes of some Bucerotiformes, Caprimulgiformes, Cariamiformes, Charadriiformes, Psittaciformes, and Strigiformes species (Sammler et al. 2011; Urantowska et al. 2018), some Ciconiformes, Gaviiformes, Musophagiformes, Podicipediformes, Sphenisciformes (Urantowska et al. 2020), some Passeriformes (Formenti et al. 2021), the majority of Pelecaniformes (Gibb et al. 2007; Zhou et al. 2014) and Procellariiformes (Abbott et al. 2005; Gibb et al. 2007; Lounsbury et al. 2015) as well as all representatives of Gruidae (Akiyama et al. 2017) and Suliformes (Morris-Pocock et al. 2010; Gibb et al. 2013; Rodrigues et al. 2017; Zhang et al. 2017).

Subsequent degeneration of the fully duplicated region due to pseudogenization or loss of some genes and/or control region led to the occurrence of all other avian gene orders containing the duplicated elements (Eberhard and Wright 2016; Urantowska et al. 2018, 2020). Two of these rearrangements were identified in Accipitriformes. *Pandion haliaetus* contains the mitogenome, in which only the control region is duplicated (Gibb et al. 2007), but single genes of *tRNA-Pro*,

ND6, and *tRNA-Glu* are placed in other order (GO-II in fig. 1). Most probably, such rearrangement could result from the GO-FD gene order degeneration due to the loss of the first copies of *tRNA-Pro*, *ND6*, and *tRNA-Glu* genes as well as the second copy of cytochrome b and *tRNA-Thr* gene. Within all other known 35 Accipitriformes mitogenomes, another gene order was identified (GO-IV in fig. 1), in which the degeneration is more advanced, because the second copy of the control region is significantly shortened and/or diverged at the sequence level in comparison to the first one. The GO-II gene order was also reported in some representatives of Coraciiformes (Huang et al. 2016), Passeriformes (Mackiewicz et al. 2019), Piciformes (Gibb et al. 2007), Psittaciformes (Guan et al. 2016), and Strigiformes (Hanna et al. 2017), whereas the GO-IV rearrangement was identified in the mitogenomes of some Charadriiformes (Grealy et al. 2019), Falconiformes (Mindell et al. 1998; Gibb et al. 2007; Mackiewicz et al. 2019), and Passeriformes (Mackiewicz et al. 2019), most Piciformes (Tamashiro et al. 2019) as well as all Cuculiformes (Pratt et al. 2009; Wang et al. 2016).

The presence of identical rearrangements containing the duplicated elements in different avian lineages strongly suggests their common evolutionary origin from an ancestral mitogenomic variant because their independent origin would require too many molecular events leading to the same gene order (Boore and Brown 1998; Rokas and Holland 2000; Boore 2006; Boore and Fuerstenberg 2008). In agreement with that the recent analyses demonstrated that a duplication was present not only in the common ancestor of Psittaciformes (Urantowska et al. 2018) and that of Psittaciformes, Passeriformes, and Falconiformes (Mackiewicz et al. 2019) but also Palaeognathae or possibly early diverging lineages of Neoaves and even all birds (Urantowska et al. 2020).

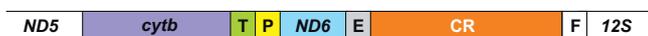
The fact that mitogenomes of almost all members of Accipitriformes contain duplicated fragments (Roques et al. 2004; Gibb et al. 2007; Qin et al. 2013; Jiang et al. 2015; Liu et al. 2017) strongly suggests that the duplicated gene order could be the ancestral rearrangement also for this groups. Unfortunately, up to now, the lack of duplicated fragments within the mitogenome of *Sagittarius serpentarius*, the early diverged lineage of Accipitriformes (Mahmood et al. 2014), makes this thesis less probable. Similarly, the single gene order was also found in *Cathartes aura* (Slack et al. 2007) belonging to Cathartiformes, a sister order to Accipitriformes. These results suggest that the mitogenomic duplication has not been inherited from the common ancestor of Accipitriformes but occurred later in the evolution of this order.

However, the lack of duplication in these mitogenomes may result from the omission of identical repeats due to an inappropriate PCR strategy, insufficient sequencing methods, or incorrect genome assembly. Reanalyses involving the appropriate approaches revealed the presence of duplicated

A Typical vertebrate gene order



B Typical avian gene order



C Ancestral duplicated gene order



D Rearrangements which evolved in Cathartiformes, Accipitriformes and Strigiformes

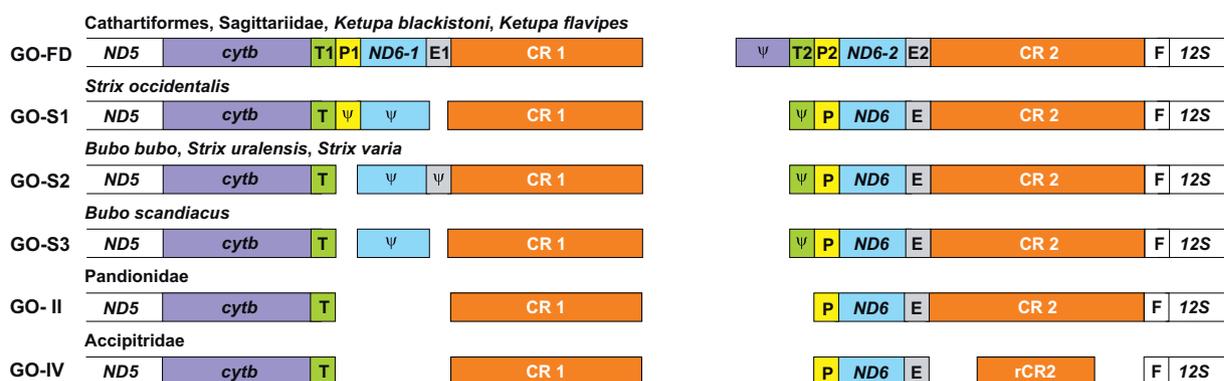


FIG. 1.—The comparison of mitochondrial gene orders between *ND5* and *12S rRNA* for a typical vertebrate gene order (A), a typical avian gene order (B), an ancestral duplicated gene order assuming the tandem duplication of the *cytb* to CR segment (C), and rearrangements observed in Cathartiformes, Accipitriformes, and Strigiformes (D) and including the most fully duplicated avian gene order (GO-FD) as well as gradually degenerated rearrangements in Accipitriformes (GO-II, GO-IV) and Strigiformes (GO-S1, S2, and S3). *ND5*, NADH dehydrogenase subunit 5 gene; *cytb*, cytochrome b gene; T, tRNA gene for threonine; P, tRNA gene for proline; *ND6*, NADH dehydrogenase subunit 6; E, tRNA gene for glutamic acid; CR, control region; F, tRNA gene for phenylalanine; *12S*, *12S rRNA* gene. Pseudogenes are marked by ψ and colored correspondingly to their functional gene copy.

gene orders in several mitogenomes from various avian groups, which were previously annotated without the duplication (Gibb et al. 2007, 2015; Akiyama et al. 2017; Urantowka et al. 2018). Therefore, a diligent searching for potentially unidentified duplications is crucial to establish the true frequency of their occurrence. It is necessary to infer ancestral avian rearrangements and understand the evolution of avian mitochondrial genome. Results of these studies are interesting in the general context of mitogenome structure and organization, which co-occurred with the evolution of multicellular animals (Lavrov 2007, 2011).

Re-examination of gene order within mitogenomes of *Sagittarius serpentarius* and representatives of Cathartiformes plays a pivotal role in inferring the mitogenomic rearrangement ancestral for Accipitriformes as well as the whole Accipitrimorphae. It cannot be ruled out that the potential presence of identical repeats within *Sagittarius serpentarius* and *Cathartes aura* mitogenomes prevented their detection with an inadequate PCR strategy. For this reason, to look for the potentially omitted duplications, we used a PCR strategy that allows amplifying the fragment between two control regions and complete mitogenomes. A comparison of the new data enabled to elucidated the complex evolution of their mitogenomes in terms of duplication as well as obtain a well-resolved phylogeny of Accipitriformes and Cathartiformes.

Results

Phylogenetic Relationships within Accipitrimorphae Based on Complete Mitogenomes

The inclusion of complete mitogenomes enabled us to obtain well-resolved phylogenetic relationships within Accipitrimorphae (fig. 2). All deep nodes received the highest support in all applied methods. *Sagittarius serpentarius* turned out the earliest diverged lineage of Accipitriformes and next, *Pandion haliaetus* branched off. These species are sister to Accipitridae whose basal lineage is *Pernis ptilorhynchus*. *Spilornis cheela* is clustered significantly with representatives of monophyletic subfamily Aegypiinae. Aquilinae is also monophyletic and related to the clade that consists of two subclades: 1) representatives of Milvinae and Haliaeetinae grouped with Buteoninae and 2) Circinae placed within Accipitridae. The latter subfamily is represented by *Accipiter*, which is not monophyletic because its two species, that is, *Accipiter gentilis* and *Accipiter nisus*, are phylogenetically closer to the species of *Circus* than to other three species of *Accipiter*.

Identification of Duplicated Gene Order in Mitogenomes of Cathartiformes and *Sagittarius serpentarius*

Using an appropriate PCR strategy (fig. 3), we obtained the diagnostic fragments between the first (CR1) and the second

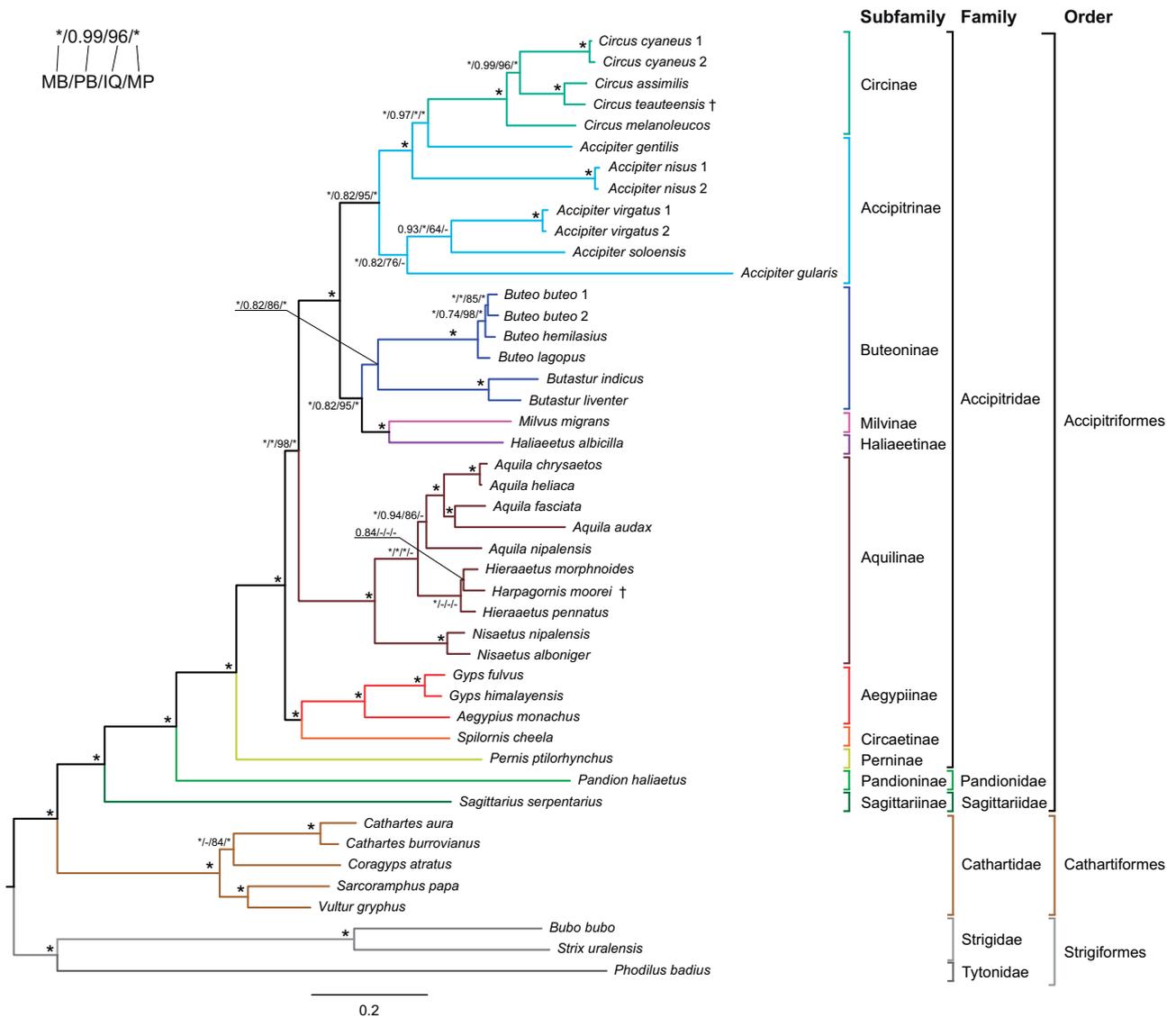


FIG. 2.—Phylogenetic tree obtained in MrBayes showing relationships between Cathartiformes and Accipitriformes with Strigiformes used as an outgroup. The values at nodes, in the following order MB/PB/IQ/MP, indicate: posterior probabilities found in MrBayes (MB) and PhyloBayes (PB) as well as bootstrap percentages in IQ-TREE (IQ) and support in Shimodara–Hasegawa-like approximate likelihood ratio test calculated in morePhyML (MP). The asterisk “*” indicates the maximal possible support value. The posterior probabilities <0.5 and the percentages <50% were marked by a dash “-.” The number after species name indicates the number of individual.

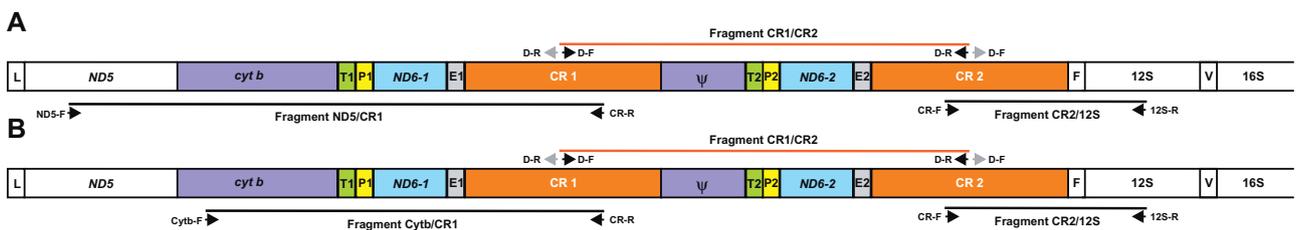


FIG. 3.—Strategy used in this study for identification of gene orders within duplicated regions in the mitogenomes of *Cathartes aura* (A), *Cathartes burrovianus* (B), *Coragyps atratus* (B), *Sagittarius serpentarius* (B), *Sarcoramphus papa* (B), and *Vultur gryphus* (B). L, tRNA for leucine; ND5, NADH dehydrogenase subunit 5; *cytb*, cytochrome b; T, tRNA for threonine; P, tRNA for proline; ND6, NADH dehydrogenase subunit 6; E, tRNA for glutamic acid; CR, control region; F, tRNA for phenylalanine; 12S, 12S rRNA; V, tRNA for valine; 16S, 16S rRNA. ND5-F, Cytb-F, CR-R, D-F, D-R, CR-F, 12S-R: primers that were used for amplification of three overlapping mitogenomic fragments.

Table 1

Species Analyzed in This Study in Terms of Duplicated Regions

Species	Source (number of individuals) ^a	Fragment	Length (bp)	Accession Number
<i>Sagittarius serpentarius</i>	ZOO KRAL (1); ZOO BER (2)	Cytb/12S rRNA	7,108	MN720443
<i>Cathartes aura</i>	ZOO GDA (1)	ND5/12S rRNA	7,969	MN629891
<i>Cathartes burrovianus</i>	ZOO PRA (1); ZOO BER (2)	Complete mitogenome	19,285	MN720441
<i>Coragyps atratus</i>	ZOO PLZ (1); ZOO FRA (2)	Complete mitogenome	19,329	MN720440
<i>Sarcoramphus papa</i>	ZOO PRA (1); ZOO PLZ (1); ZOO BER (2)	Complete mitogenome	19,302	MN720442
<i>Vultur gryphus</i>	ZOO WAW (1)	Complete mitogenome	19,287	MN720444

NOTE.—Species whose mitochondrial genomes were newly sequenced in this study are bolded.

^aSources of samples: ZOO BER, Tierpark Berlin-Friedrichsfelde GmbH; ZOO FRA, Zoo de la Boissière du Doré; ZOO GDA, Zoological Garden in Gdańsk; ZOO KRAL, Zoological Garden in Dvůr Králové; ZOO PLZ, Zoological Garden in Plzeň; ZOO PRA, Zoological Garden in Prague; ZOO WAW, Zoological Garden in Warsaw.

(CR2) control regions for all or most of nine reactions (supplementary fig. S1, Supplementary Material online). Sequencing and annotation of these amplicons revealed the presence of *tRNA-Thr*, *tRNA-Pro*, *ND6*, and *tRNA-Glu* genes between two control regions in all cases. Furthermore, in the region located between CR1 and *tRNA-Thr*, we found a sequence identical or highly similar to the 3' end of the full-length *cytb* in all examined taxa. These truncated versions deprived of the 5' region are most likely pseudogenes. Additional PCR reactions completed the missing parts of CRs (fig. 3 and supplementary table S1, Supplementary Material online). The mitogenomic fragments containing the duplicated regions were obtained by assembling three overlapping fragments (4, 5, and 6 in supplementary table S1, Supplementary Material online). In all cases, the same GO-FD gene order was found (fig. 1D). Additional PCR reactions for fragments L, 1, 2, and 3 (supplementary tables S1 and S2, Supplementary Material online) were run to complete the missing parts of mitogenomes from *Cathartes burrovianus*, *Coragyps atratus*, *Sarcoramphus papa*, and *Vultur gryphus*. Specific and efficient products were obtained for three or nine pairs of primers depending of the species (supplementary fig. S2, Supplementary Material online). After assembling four or six overlapping fragments (supplementary table S1, Supplementary Material online), the length of complete mitogenomes varied from 19,285 to 19,329 bp (table 1).

Duplicated Regions in Mitogenomes of Cathartiformes and Accipitriformes in Comparison to Strigiformes

The fully duplicated region (GO-FD) found in *Sagittarius serpentarius* and five species of Cathartiformes is characterized by very high similarity between paralogous sequences, that is, copies found within the same mitogenome (fig. 1). The second copies of *tRNA-Thr*, *tRNA-Pro*, *ND6*, and *tRNA-Glu* are identical with their corresponding first copies (table 2 and supplementary fig. S3, Supplementary Material online). Although the paralogous *cytb* sequences are also identical or differ by at most one nucleotide in *Cathartes burrovianus*, the second copies are substantially truncated from their originals by 595–861 bp, leaving only the 3' end (table 2). These

incomplete copies constitute less than 48% of the full-length *cytb*, which means that the functional domain is substantially truncated. Moreover, the degenerated copies do not represent the full-length open reading frames. Therefore, they are most likely pseudogenes. The high sequence similarity of the truncated copy to the full version can indicate that the duplication and concerted evolution occurred quite recently. Two control regions also show high sequence identity, from 92% to 97% (table 2) and the low percent of gaps, from 19% to 23% (supplementary table S3, Supplementary Material online). The differences are located at their 5' and 3' ends, except for *Sagittarius serpentarius*, whose control regions differ only at their 3' ends (supplementary fig. S3, Supplementary Material online). The greater variation at the ends of CRs can result from the presence of microsatellites and repeated sequences.

The GO-II gene order found in *Pandion haliaetus* is deprived of the first copies of *tRNA-Pro*, *ND6*, and *tRNA-Glu* but retains their second copies (fig. 1). In turn, *tRNA-Thr* is represented by the first copy and its second copy was lost together with the adjacent *cytb* pseudogene. The control regions are still very similar with 98% of identical nucleotides and only 18% of gaps in the global alignment (supplementary fig. S4 and table S3, Supplementary Material online). Representatives of Accipitridae show the gene order GO-IV, similar to GO-II but CR2 is much more degenerated (fig. 1). The percent of identity is from 72% to 84%, whereas the percent of gaps from 45% to 84% (supplementary fig. S4 and table S3, Supplementary Material online).

CR2s (1,262–1,330 bp) are longer than CR1s (1,066–1,092 bp) in all five examined Cathartiformes (supplementary table S4, Supplementary Material online), whereas CR1 in *Sagittarius serpentarius* is 236 bp longer than CR2, because it contains a 77-bp sequence repeated seven times at its 3' end. CRs of *Pandion haliaetus* are comparable in length, CR2 is only 33 bp longer. In the case of gene order GO-IV, CR2s are usually shorter (334–2,606 bp) than CR1s (1,144–3,571 bp), especially in the species of *Circus* and *Accipiter* (supplementary table S4, Supplementary Material online). The length of 5' spacer proceeding poly-C motif of CR1 is smaller (24–25 bp)

Table 2 Comparison of Genes and Control Regions (CR) in the First and Second Copies of Duplicated Region in the Mitogenomic Sequences Obtained in This Study for Six Raptor Taxa

Species	Copy	Length (bp)		Number/Percent of Residues Identical between Two Copies of									
		cytb	CR	tRNA-Thr	tRNA-Pro	ND6	tRNA-Glu	CR	cytb	tRNA-Thr	tRNA-Pro	ND6	tRNA-Glu
<i>Cathartes aura</i>	1st	1,144	69	70	522	73	1,092	534/100	69/100	70/100	522/100	73/100	1,016/96.0
	2nd	534	69	70	522	73	1,321						
<i>Cathartes burrovianus</i>	1st	1,144	69	70	522	73	1,092	536/99.8	69/100	70/100	522/100	73/100	1,017/96.8
	2nd	537	69	70	522	73	1,262						
<i>Coragyps atratus</i>	1st	1,144	69	70	513	73	1,090	546/100	69/100	70/100	513/100	73/100	1,035/96.9
	2nd	546	69	70	513	73	1,330						
<i>Sarcoramphus papa</i>	1st	1,144	70	70	522	74	1,083	548/100	70/100	70/100	522/100	74/100	1,030/97.4
	2nd	548	70	70	522	74	1,288						
<i>Vultur gryphus</i>	1st	1,144	69	70	522	74	1,066	548/100	69/100	70/100	522/100	74/100	1,007/96.6
	2nd	549	69	70	522	74	1,299						
<i>Sagittarius serpentarius</i>	1st	1,064 ^a	71	70	522	71	1,969	203/100	71/100	70/100	522/100	71/100	1,483/92.2
	2nd	203	71	70	522	71	1,733						

^aNot full version of this gene was obtained.

in GO-FD and the less reduced GO-II gene orders than in the much degenerated order GO-IV (53–93 bp) (supplementary table S4, Supplementary Material online). The poly-C motif at the CR2 can be identified only in GO-FD and the less reduced GO-II and *Pernis ptilorhynchus* with the GO-IV order.

When we compare the mitogenomic gene orders in Cathartiformes and Accipitriformes with those in Strigiformes, we can notice complementary stages in the gradual degeneration of duplications (fig. 1). The fully duplicated region (GO-FD) can be found in two *Ketupa* species (supplementary fig. S5 and table S5, Supplementary Material online). *Strix occidentalis* has a gene order (GO-S1) already without *tRNA-Glu1* and *cytb-2*, whereas *tRNA-Pro1*, *ND6-1*, and *tRNA-Thr2* are pseudogenized (supplementary fig. S5 and table S5, Supplementary Material online). GO-S2, present in *Bubo bubo*, *Strix uralensis*, and *Strix varia*, is similar to GO-S1 but lost *tRNA-Pro1* and maintained *tRNA-Glu1* pseudogenized. *Bubo scandiacus* is deprived of both these genes (GO-S3) (supplementary fig. S5 and table S5, Supplementary Material online). More advanced steps in the degeneration are represented by GO-II in Pandionidae and GO-IV in Accipitridae (fig. 1).

Similar to *Sagittarius serpentarius* and Cathartiformes with GO-FD, *Ketupa blackstoni* and *Ketupa flavipes* have also the same or almost the same length of the corresponding duplicated genes of *tRNA-Thr*, *tRNA-Pro*, *ND6*, and *tRNA-Glu*, which show high sequence identity of 97–100% without any gaps. The second copies of *cytb* are also almost identical, but are shorter than the original ones by 942 and 949 bp, respectively. In gene orders GO-S1, S2, and S3, the pseudogenization of *tRNA-Thr*, *tRNA-Pro*, *ND6*, and *tRNA-Glu* is more advanced. Their sequences were significantly truncated, *tRNAs* by 3–62 bp and *ND6* by 420–497 bp, or filled with repeats and other insertions, for example, *tRNA-Thr* in *Strix occidentalis* and *Strix varia*. As a result, their identity dropped to 67–90%.

Like nondegenerated gene orders in Accipitrimorphae, two copies of most Strigiformes CRs are also characterized by high sequence identity (93–98%) and a small percent of gaps (16–29%). Only *Bubo bubo* showed a deviated value of nucleotide identity, that is, 85%. In *Ketupa blackstoni* and *Ketupa flavipes*, CR1s are longer than CR2s by 277 and 299 bp, whereas in other owl mitogenomes CR2s are characterized by a greater length by 146–206 bp.

Distribution of Rearrangement Types in the Phylogenetic Tree of Accipitrimorphae

We mapped the presence of three mitogenomic gene orders onto the phylogenetic tree of Accipitrimorphae (fig. 4). The maximum parsimony and likelihood methods inferred that the fully duplicated gene order (GO-FD) was the plesiomorphic state in the common ancestor of Cathartiformes and Accipitriformes with the probability of $P = 0.998$ as well as

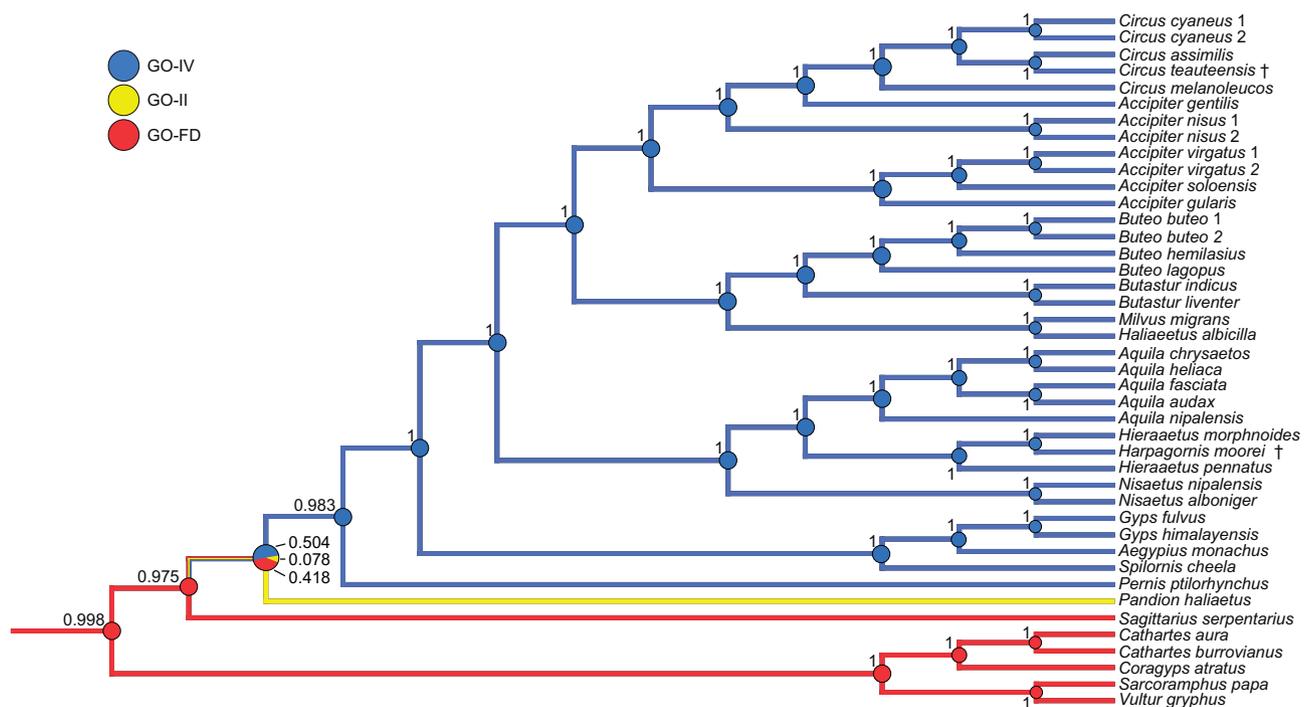


FIG. 4.—The maximum parsimony and maximum likelihood reconstruction of ancestral states as well as mapping of three mitochondrial gene orders (GO-FD, GO-II, GO-IV, shown in fig. 1) onto the phylogenetic relationships of Cathartiformes and Accipitriformes. The values and area of colors at nodes correspond to the probability of the given ancestral state. The number after species name indicates the number of individual.

Accipitriformes with $P=0.975$. This character was maintained both in modern Cathartiformes and *Sagittarius serpentarius* lineage. The state of the common ancestor of Pandioninae and Accipitridae is not unambiguously resolved (fig. 4). It could have GO-FD or degenerated GO-IV gene order. However, other nodes have assigned states with a much higher probability. The common ancestor of Accipitridae had the degenerated gene order GO-IV with $P=0.983$, which was inherited to its descendants. In *Pandion haliaetus*, a version with the unreduced control region, that is, GO-II, evolved.

Clustering Sequences of Duplicated Control Regions

The mapping of mitogenomic gene orders onto the phylogenetic relationships gives only a general picture of evolution of duplicated regions, which in fact could be more complex. Therefore, we subjected duplicated copies of CRs to more detailed investigations. Their comparison revealed the presence of several clusters in the 2D space of sequence similarity (fig. 5). Most CRs, both the first and second copies, are grouped together but there are CR2s from taxonomically related genera *Accipiter* and *Circus* as well as *Pernis ptilorhynchus* that are far away from the main group in the space. This indicates a very high divergence of these sequences.

In the main group extracted in panel 1 in figure 5, we can recognize six clusters reflecting phylogenetic relationships of their members. The most distant cluster 1 in this group

contains CR2s from closely related subfamilies Circaetinae and Aegypiinae as well Buteoninae and Haliaeetinae. CR2 of the representative of Milvinae, also phylogenetically related to the latter, is closely located to the cluster 1. All Aquilinae CR2s are grouped in the cluster 2 and relatively strongly connected with the cluster 3 including CR1s also from the same subfamily. In turn, CR1s from five phylogenetically related subfamilies, that is, Circinae, Accipitrinae, Buteoninae, Milvinae, and Haliaeetinae are grouped in the neighboring cluster 4. The cluster 5 consists of both CR1s and CR2s from Cathartiformes. Two copies of CRs from Strigiformes create also a separate cluster 6. CR1 from *Spilornis cheela* can be also classified into the cluster 3, whereas CR1s from Aegypiinae and *Pernis ptilorhynchus* are in the surroundings of the clusters 3 and 4. The pairs of two CRs from early diverged lineages of Accipitriformes, *Pandion haliaetus* and *Sagittarius serpentarius*, are placed close to the clusters 3 and 5.

The strong connections between CR1 and CR2 from the same species, that is, *Pandion haliaetus* and *Sagittarius serpentarius*, or closely related taxa in Cathartiformes and Strigiformes could indicate a quite recent duplication or homogenization of these regions. On the other hand, the distant location of CR1s and CR2s from higher taxonomic units, for example, Aquilinae, suggests that the duplication or homogenization occurred before diversification of these groups.

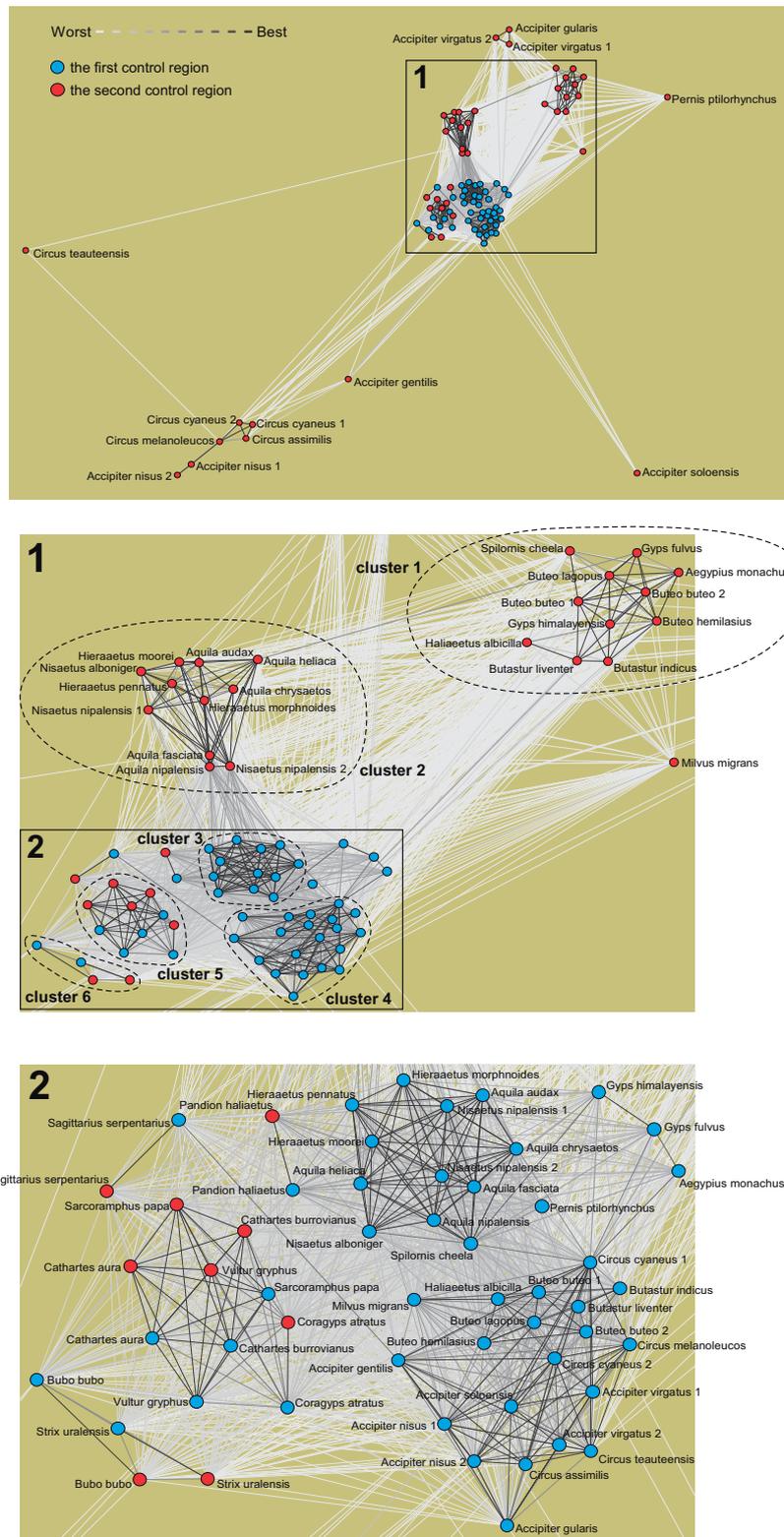


Fig. 5.—Cluster analysis of 45 pairs of control regions sequences (the first and the second copies) from representatives of Accipitriformes, Cathartiformes, and Strigiformes made in CLANS software (Frickey and Lupas 2004). The analyzed sequences are represented by vertices connected by edges reflecting attractive forces proportional to the negative logarithm of E-value calculated for the high scoring segment pairs (HSPs). The gray shade intensity of the connections is proportional to these forces. The edges with E-value <0.001 are shown. The number after species name indicates the number of individual.

Phylogenetic Relationships between Duplicated Control Regions in Accipitriformes

In order to obtain better resolved relationships between the CRs, we conducted more detailed phylogenetic studies based on their sequences. Since many CR2s were characterized by a high divergence, their unrelated sequences could be artificially clustered in phylogenetic trees due to the long-branch attraction (LBA) artifact (Felsenstein 1978). Therefore, we studied separately the individual CR2s or those from closely related taxa, for example, genera and subfamilies, in phylogenetic analyses with CR1s in various combinations of taxa. All data sets included also other CR2s but only those that were almost identical to their CR1s and did not show the high substitution rate. We investigated data sets on various evolutionary distances containing taxa from only Accipitriformes, as well as additionally from Cathartiformes and also Strigiformes to check the outgroup influence. Moreover, we prepared three alignment types, containing: all unmasked sites, masked low complexity and repeated regions as well as deprived of them before aligning the sequences. We also applied the covarion model of nucleotide substitutions to minimize LBA. In consequence, the combination of various taxa and selected sites provided 76 sequence alignments. The selected data sets and groups are included in [supplementary table S6, Supplementary Material](#) online and presented in the produced trees in [supplementary figures S6–S8, Supplementary Material](#) online.

The position of CR2s in the inferred phylogenetic trees with CR1s should depend on the evolutionary time when the duplication of the control region occurred. If orthologous CR copies are grouped together, that is, CR2s from a group of species create a clade that is clustered with CR1s from the same species set, we could infer that the duplication or homogenization occurred before the divergence of these species. Alternatively, if two paralogous CR copies from the same species are grouped together, we could deduce that the duplication or homogenization happened quite recently in the lineage of a given species. The second case could also take place when the duplicated regions were subjected to homogenization via concerted evolution (Kumazawa et al. 1996, 1998; Eberhard et al. 2001; Ogoh and Ohmiya 2007; Tatarenkov and Avise 2007; Akiyama et al. 2017). Due to this process, sequences of duplicated regions became very similar or identical, which imitates a recent duplication event.

The analysis of phylogenetic trees revealed these two types of expected positions of CR2s. Generally, CR1 sequences follow the phylogenetic relationships between the analyzed taxa inferred on the complete mitochondrial genomes. We summarized the results in the form of consensus cladograms obtained from the individual phylogenetic trees. [Figures 6 and 7](#) show selected consensus cladograms, whereas all such cladograms are included in [supplementary figure S6, Supplementary Material](#) online. Selected individual

phylograms with branch lengths are presented in [supplementary figure S7, Supplementary Material](#) online and the corresponding individual cladograms with support values are in [supplementary figure S8, Supplementary Material](#) online.

In agreement with the clustering analyses, the two copies of CRs from individual species, that is, *Pandion haliaetus* and *Sagittarius serpentarius* as well as the representatives of Strigiformes are significantly grouped together, obtaining often the maximal support in all approaches ([figs. 6 and 7; supplementary figs. S6 and S8, Supplementary Material](#) online). Tree branches leading to these CR sequences are of a similar length and are quite short ([supplementary fig. S7, Supplementary Material](#) online). Two CR sequences of *Spilornis cheela* are also clustered together in almost all trees by four methods ([fig. 6](#)). However, the branch of CR2 sequence is very long, indicating a large number of substitutions ([supplementary fig. S7, Supplementary Material](#) online).

In contrast to that, CR2s of individual species *Aegypius monachus* and *Gyps fulvus* are not grouped with their paralogous copies CR1s but are located mostly after the divergence of *Pernis ptilorhynchus* CR and before the radiation of other Accipitridae subfamilies ([fig. 6 and supplementary figs. S6 and S7, Supplementary Material](#) online). When CR2s of both *Aegypius monachus* and *Gyps fulvus* are included in the analysis, they are clustered together with the maximal support in all methods and occupy the same phylogenetic position in the trees as the individual sequences. The CR2 sequence from *Pernis ptilorhynchus* was placed after the separation of *Sagittarius serpentarius* and before the divergence of *Pandion haliaetus* in one tree with a poor support ([supplementary fig. S8, Supplementary Material](#) online).

CR2s of Aquilinae are joined into one cluster, which obtained the maximal support in all approaches and was sister to the clade of CR1s also from Aquilinae in almost all trees ([fig. 7 and supplementary figs. S6 and S8, Supplementary Material](#) online). We also found that CR2s sequences from the members of closely related subfamilies, that is, Circinae, Accipitrinae, Buteoninae, Milvinae, and Haliaeetinae, create a very strongly supported group, which is linked with the cluster of CR1 counterparts from the same subfamilies in the vast majority of trees ([fig. 7 and supplementary figs. S6 and S8, Supplementary Material](#) online). This grouping obtained the maximal support values ([fig. 7](#)). To avoid a potential long-branch attraction of the highly diverged CR2s sequences from these subfamilies, we studied separately groups of the subfamilies, which are closely related, that is, Accipitrinae + Circinae as well as Buteoninae, Buteoninae + Milvinae, and Buteoninae + Milvinae + Haliaeetinae. In all cases, these groups are clustered into clear significant clades. Like in the analysis including CR2s from representative of all these families, the clades of these groups are joined with the cluster of CR1 sequences from Circinae, Accipitrinae, Buteoninae, Milvinae, and Haliaeetinae in almost all trees with the maximal

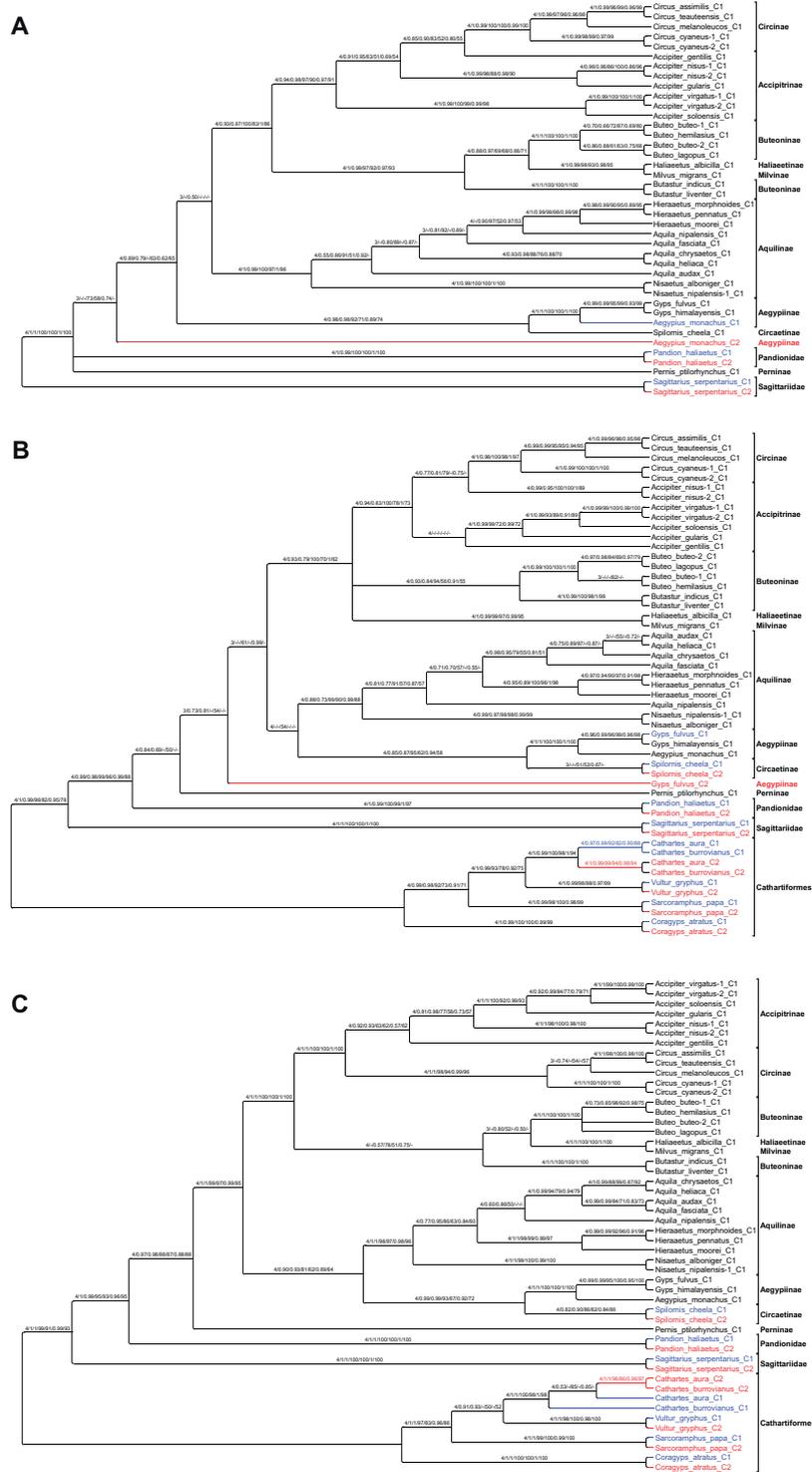


Fig. 6.—Consensus cladograms of phylogenetic trees inferred in four programs, MrBayes, PhyloBayes, IQ-TREE, and (more)PhyML, including the second control region copies of *Aegypius monachus* (A), *Gyps fulvus* (B), and *Spilornis cheela* (B and C). Pairs of CRs from the same species are colored, whereas CRs without the second copy in the tree are in black. The blue and red colors indicate the corresponding first and second copies of CR, respectively. The taxa names are in the format Genus_species-X_CY, where X is the individual number (if present) and Y is the number of control region, that is, 1 or 2. The values at nodes, in the following order N/MB/PB/SH-I/VP-I/SH-P/BP-P, indicate the number of trees containing a given node (N), posterior probabilities found in MrBayes (MB) and PhyloBayes (PB), as well as SH-aLRT and nonparametric bootstrap support values calculated in IQ-TREE (SH-I and BP-I) and (more)PhyML (SH-P and BP-P). The posterior probabilities <0.5 and the percentages <50% were indicated by a dash “-.” SH-aLRT means approximate likelihood ratio test based on Shimodara–Hasegawa procedure. See [supplementary table S6, Supplementary Material](#) online, for details about the data sets.

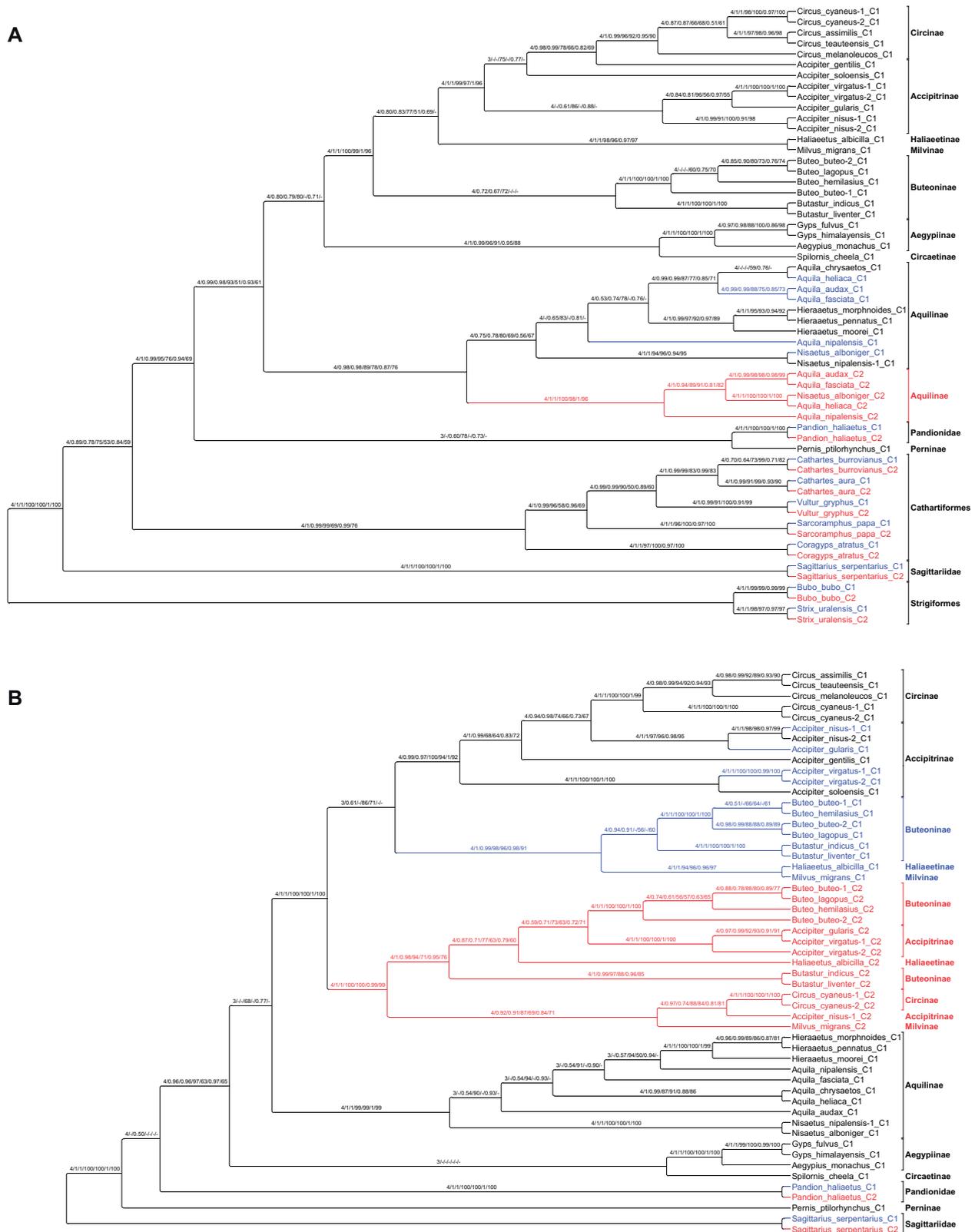


Fig. 7.—Consensus cladograms of phylogenetic trees inferred in four programs, MrBayes, PhyloBayes, IQ-TREE, and (more)PhyML, including the second control region copies of representatives from Aquilinae (A) as well as Circinae, Accipitrinae, Buteoninae, Milvinae, and Haliaeetinae (B). For other explanations, see figure 6.

support values (supplementary figs. S6 and S8, Supplementary Material online).

In the case of all CR2s that are not grouped with their corresponding first copies from the same species, the length of their branches in the phylogenetic trees is much larger than that of the corresponding CR1s (supplementary fig. S7, Supplementary Material online). It indicates that the CR2 sequences were subjected to a much higher accumulation of substitutions than their CR1 paralogs.

Testing the Alternative Phylogenetic Positions of CR2 Sequences

Despite the fact that many of the inferred phylogenetic positions of CR2 sequences were strongly supported, we also tested the alternative groupings of the fast evolving CR2 sequences with CR1 copies. We considered that these CR2 sequences are grouped with more deep branches, that is, closer to the tree root, which suggests that the control regions duplicated earlier than observed in the phylogenetic trees. We also assumed that some of CR2s are clustered with later evolving lineages or are directly grouped with their paralogous CR1 copies, which indicate their more recent duplication or homogenization via concerted evolution. In figure 8, we gathered the results of alternative topology testing and the most often positions of CR2 sequences in the phylogenetic trees. Supplementary figure S9, Supplementary Material online, includes all testing positions and results of tree topology tests as well as the observed phylogenetic positions of CR2s for individual species and groups of Accipitriformes. These positions turned out highly probable in comparison to the alternative placements. In three cases (*Pernis ptilorhynchus*, *Aegypius monachus* + *Gyps fulvus*, *Aquilinae*), the number of tree topology tests accepting these positions was greater than rejecting them, whereas in other six instances (*Spilornis cheela*, *Aegypius monachus*, *Gyps fulvus* as well as combinations of subfamilies *Circinae*, *Accipitrinae*, *Buteoninae*, *Milvinae*, and *Haliaeetinae*) no tests rejected the best topologies (fig. 8).

Phylogenetic Relationships between Duplicated Control Regions in Cathartiformes

Paralogous CRs, that is, CR1 and CR2, from the same species of Cathartiformes, such as *Coragyps atratus*, *Sarcoramphus papa*, and *Vultur gryphus*, are significantly grouped together in all four methods applied on three type of sequences (supplementary fig. S10, Supplementary Material online). However, CR sequences from *Cathartes aura* and *Cathartes burrovianus* show a variable phylogenetic pattern. Depending on the method and masking repeats in the alignment, the paralogous CRs from these species are clustered together or they are separated, whereas orthologous CRs are grouped together.

These results suggest that various parts of alignment show different evolutionary histories of CRs in *Cathartes*. To identify

individual sites in the alignment supporting one of the two tree topologies, we calculated log-likelihood values for the individual sites for these trees (supplementary fig. S11, Supplementary Material online). The analysis revealed five alignment fragments better supporting the topology t1, which segregates CR1s and CR2s of *Cathartes* species into two clades. These fragments are located approximately 55 bp from the beginning and at the end of the alignment. At the 3' end, CR2s share the same microsatellite repeats in two fragments. Six fragments conforming the topology t2, which groups paralogous CRs from the given species, are located mainly in the middle of the alignment and one is at the very beginning.

In order to check if the noticed differences in the evolution of CR sections are general and concern also other representatives of Cathartiformes, we conducted an additional analysis assuming the tree topology (t1) in which all corresponding orthologous CR copies are separated into two clades (fig. 9). The topology was compared with topology t2 grouping paralogous CRs from the same species together. Similar to the previous analysis, we found that the section with the length of 277 bp starting from the CR beginning in the 55th position of the alignment as well as the 314-bp section at the 3' end fulfils the relationships in the tree t1. In turn, the tree t2 is supported by the long middle part of CR with 819 bp and a short 48-bp fragment at the 5' end.

Variable Substitution Rate of Control Region Sequences

Phylogenetic analyses demonstrated that tree branches of individual CR copies from the same species are usually characterized by different lengths (supplementary fig. S7, Supplementary Material online). Therefore, we compared in figure 10 the differences between the lengths of branches leading to CR2s and CR1s from individual species. In almost all cases, these values are positive indicating that CR2 sequences accumulated more substitutions than CR1s. The fastest evolution revealed CR2s from *Accipiter gentilis*. Its branch lengths differ on average by approximately 3.5 substitutions per site. CR2s evolved very fast also in other representatives of Accipitrinae, that is, *Gyps fulvus* and *Butastur indicus*, because their sequences accumulated on average more than two substitutions per site than CR1s copies. The median of the difference in branch lengths from 1 to 2 can be found in *Pernis ptilorhynchus*, *Aegypius monachus*, *Nisaetus alboniger*, *Milvus migrans*, *Haliaeetus albicilla*, *Circus cyaneus*, and four representatives of Buteoninae. A slightly smaller difference but still visible is in *Spilornis cheela* and five *Aquilinae* members. The dissimilarity in the substitution rate of paralogous CRs from Strigiformes, Cathartiformes, *Sagittarius serpentarius*, and *Pandion haliaetus* is smallest but in most of these cases, the second CR copy shows still a greater substitution rate.

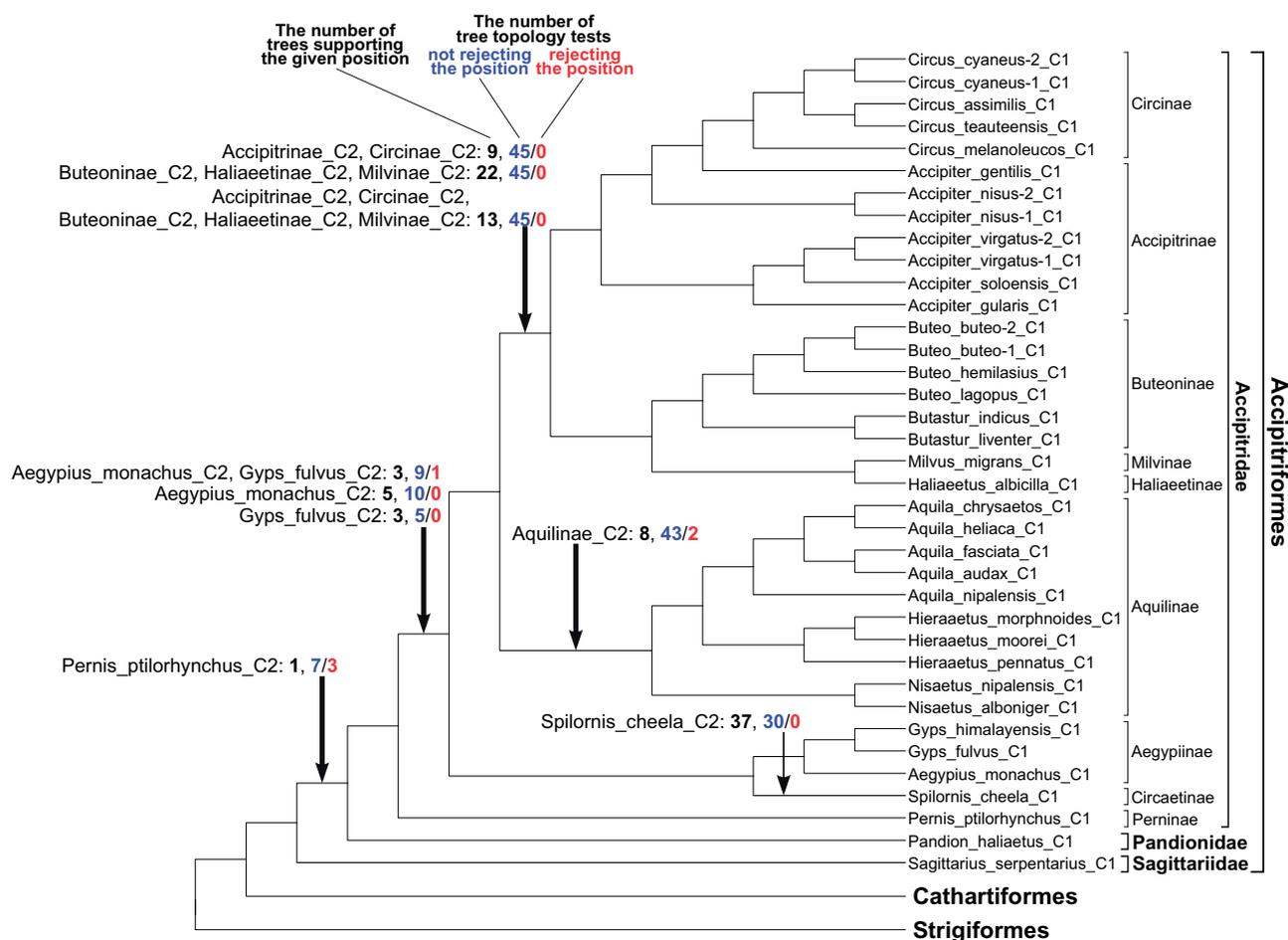


FIG. 8.—Observed phylogenetic positions of second control region copies from given species or groups of Accipitriformes. Their most frequent positions in the phylogenetic trees are indicated by arrows. The number of phylogenetic trees supporting the given position is shown in black fonts; the number of tree topology tests that did not reject the given position is shown in blue fonts; the number of tree topology tests that rejected the given position is shown in red fonts. The taxa names are in the format Genus_species-X_CY, where X is the individual number (if present) and Y is the number of control region, that is, 1 or 2.

Discussion

Phylogenetic Relationships within Accipitriformes

Using 34 mitochondrial genes, we received a robust phylogeny for Cathartiformes and Accipitriformes. Some relationships differ from those produced by other authors (Liu et al. 2017; Jiang et al. 2019). For example, they found that Aquilinae are clustered with the clade of *Aegypius monachus* + *Spilornis cheela*, whereas in our analyses Aquilinae is sister to the assemblage of Accipitrinae + Circinae and Buteoninae + Milvinae + Haliaeetinae. The other analyses included also mitogenomic data but a smaller number of sequences and not all genes, which could influence the inferred phylogenies. Our results on relationships between subfamilies of Accipitriformes and within Cathartiformes are in agreement with other authors (Jiang et al. 2015; Song et al. 2015; Johnson et al. 2016; Mindell et al. 2018; Knapp et al. 2019), but we obtained greater support values especially for

more deep nodes. We found that the genus *Accipiter* is paraphyletic, because *Circus* is nested within it. This confirms the results by Oatley et al. (2015) and Lerner et al. (2008). However, the other phylogenies were based only on two to four loci, which precluded obtaining all relationships highly supported and identifying the *Accipiter* species sister to the *Circus* clade (Oatley et al. 2015). Our results showed that *Accipiter gentilis* is most closely related to *Circus*.

Evolution of Gene Orders in Accipitriformes Mitogenomes

The phylogenetic distribution of mitogenomic gene orders and the reconstruction of ancestral states indicate that the common ancestor of Cathartiformes and Accipitriformes had the fully duplicated gene order, in which the segment of genes for cytochrome b, tRNA-Thr, tRNA-Pro, ND6, and tRNA-Glu as well as the control region were tandemly

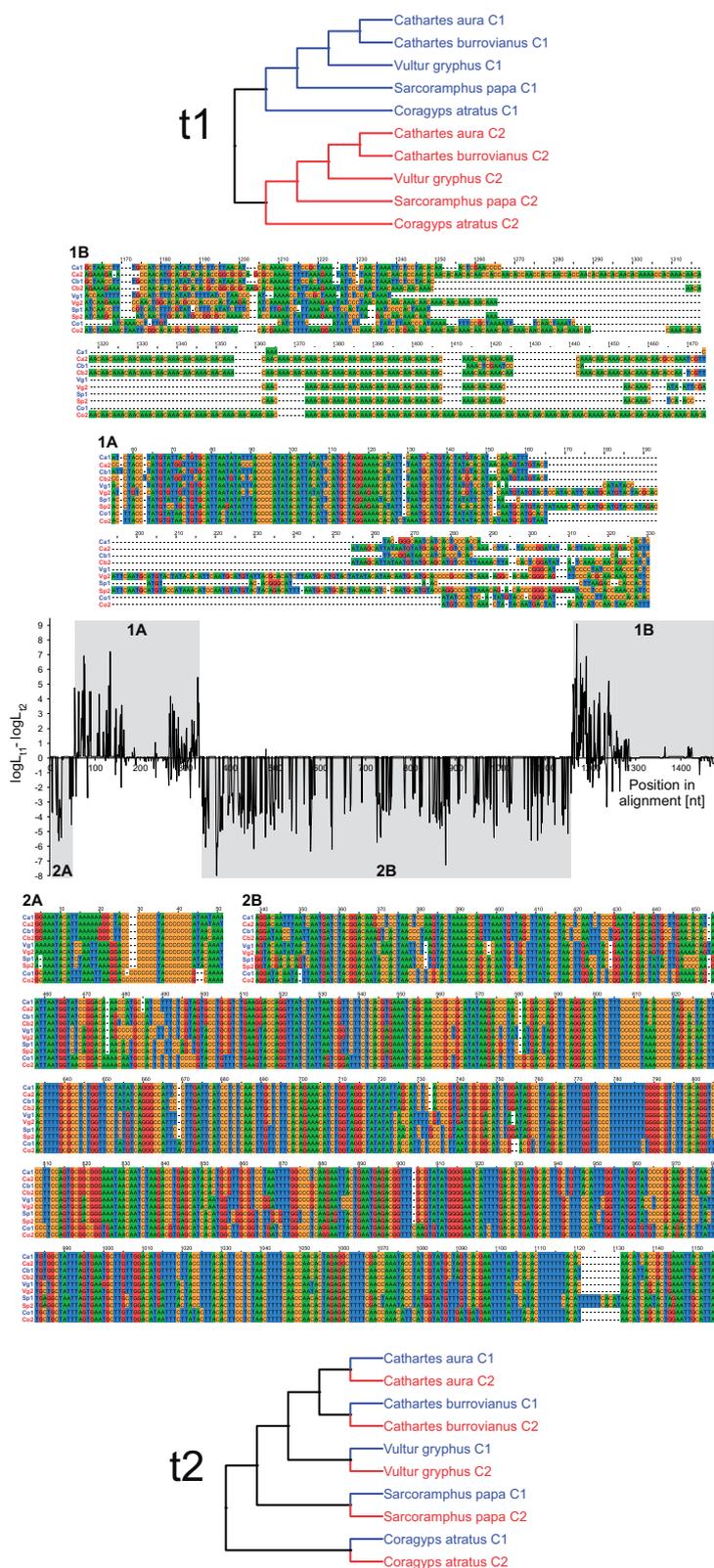


Fig. 9.—Differences between log-likelihood values for individual alignment sites of Cathartiformes control regions for two tree topologies t1 and t2 assuming different relationships between CRs (the plot inside). Above and under the plot, there are these two trees and alignment fragments supporting the given topology. The position of these fragments is marked in the plot. The labels C1 and C2 as well as blue and red colors indicate the corresponding first and second copies of CR, respectively. Ca, *Cathartes aura*; Cb, *Cathartes burrovianus*; Vg, *Vultur gryphus*; Sp, *Sarcoramphus papa*; Co, *Coragyps atratus*.

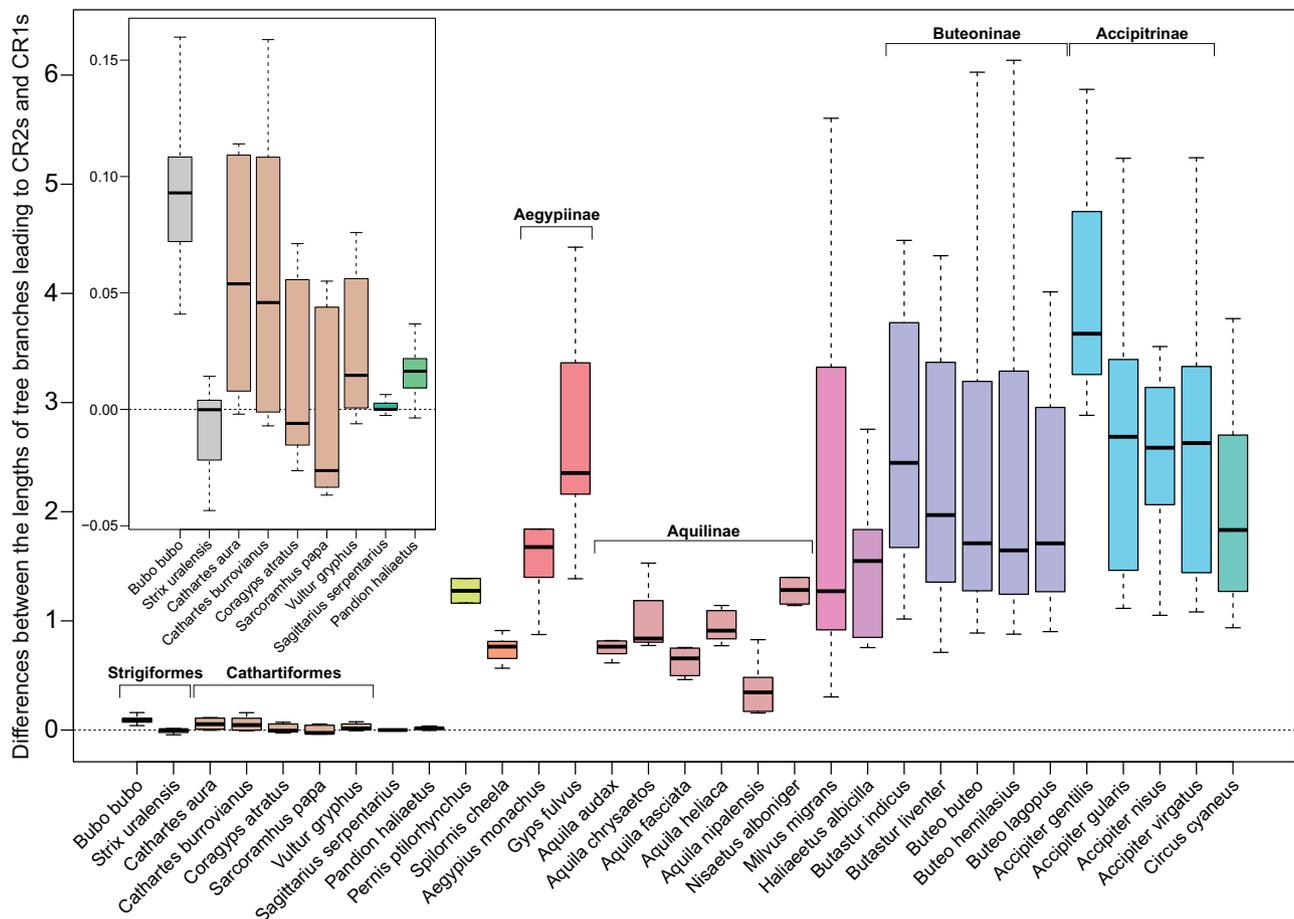


FIG. 10.—Box plots of differences between length of tree branches leading to CR2s and CR1s sequences of individual species calculated from all trees. The part of the plot for small values was enlarged and presented as an inset. The thick line indicates the median, the boxes show the quartile range and the whiskers denote the range without outliers.

deduplicated. This rearrangement is currently known as the most complete avian duplication, which was previously annotated for Bucerotiformes, Caprimulgiformes, Cariamiformes, Charadriiformes, Ciconiiformes, Gaviiformes, Gruidae, Musophagiformes, Passeriformes, Pelecaniformes, Podicipediformes, Procellariiformes, Psittaciformes, Sphenisciformes, Strigiformes, and Suliformes (Abbott et al. 2005; Gibb et al. 2007, 2013; Morris-Pocock et al. 2010; Sammler et al. 2011; Zhou et al. 2014; Lounsbury et al. 2015; Akiyama et al. 2017; Rodrigues et al. 2017; Zhang et al. 2017; Urantowka et al. 2018, 2020; Formenti et al. 2021). Since representatives of order Strigiformes, which is sister to Accipitriformes, also possess the mitogenomic duplications, we can assume that the common ancestor of these two groups contained the duplicated fragments (fig. 11). Next, in the course of evolution, the original or duplicated elements were subjected to reduction. The duplicated version of

cytochrome b was degenerated likely as first, because a pseudogene of this gene can be identified in Cathartiformes and *Sagittarius serpentarius*. Among the duplicated copies, also *tRNA-Thr* was lost but *tRNA-Pro*, *ND6*, and *tRNA-Glu* remained, whereas their original versions disappeared. The second control region shows also a tendency to degeneration due to accumulation of a greater number of substitutions than the first copy. Many CR2s gained also repeats especially at their 3' end. The decay of CR2 is the most advanced in Accipitridae.

Interestingly, similar steps of degeneration were proposed for Psittaciformes (Urantowka et al. 2018). The most degenerated gene order with the specific loss or leaving of the first or second copies resulting in GO-II and IV arrangements (fig. 1) was also identified in many representatives of other avian orders: Charadriiformes (Grealy et al. 2019), Coraciiformes (Huang et al. 2016), Cuculiformes (Pratt et al. 2009; Wang

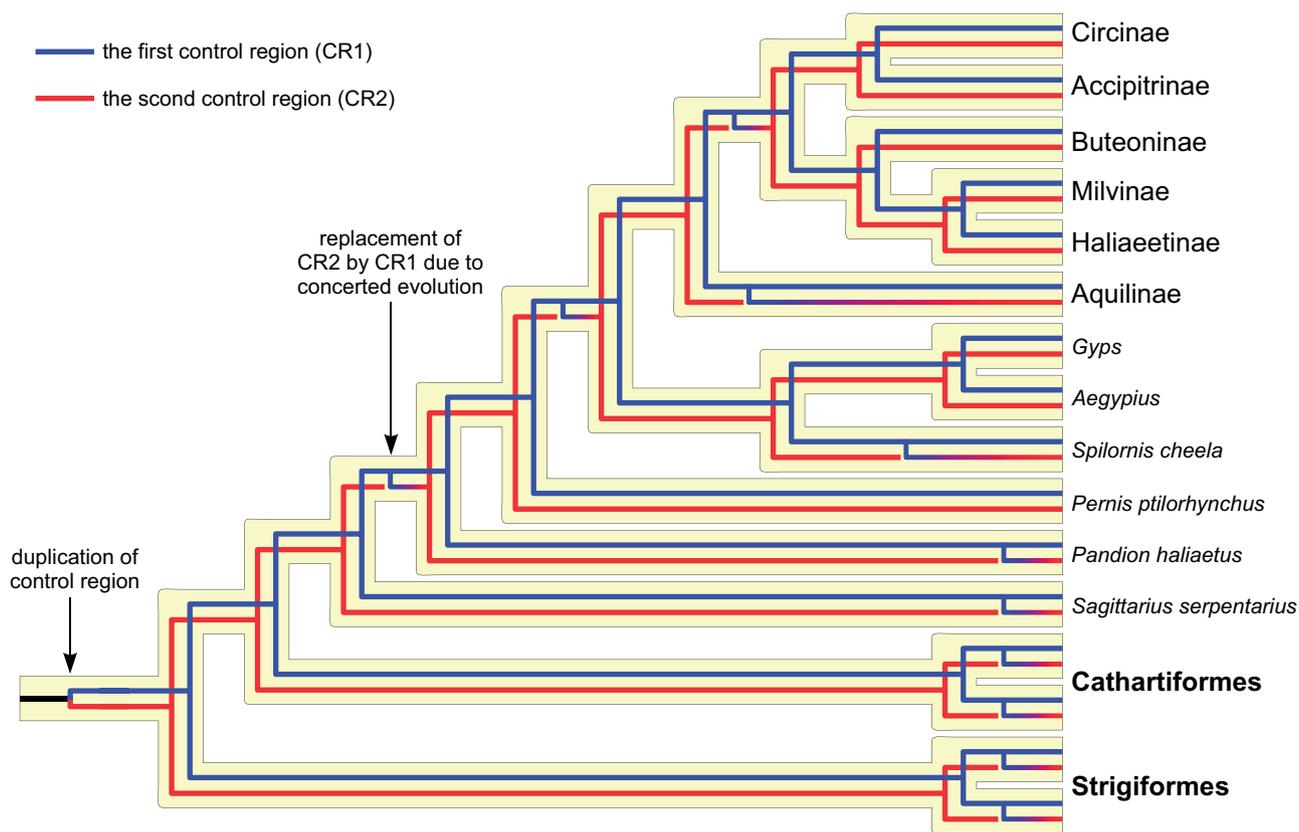


Fig. 11.—Probable evolution of control region duplications superimposed onto the phylogeny of Strigiformes, Cathartiformes, and Accipitriformes. Two CRs were indicated as blue and red lines. The tonal transition from blue to red indicates gradual homogenization of CR sequences.

et al. 2016), Falconiformes (Mindell et al. 1998; Gibb et al. 2007; Mackiewicz et al. 2019), Passeriformes (Mindell et al. 1998; Singh et al. 2008; Cooke et al. 2012; He et al. 2013; Gibb et al. 2015; Caparroz et al. 2018; Mackiewicz et al. 2019), Piciformes (Gibb et al. 2007; Tamashiro et al. 2019), and Strigiformes (Hanna et al. 2017). The predominance of this tendency is interesting, because another gene order with degeneration of the second and maintenance of the first copies of *tRNA-Pro*, *ND6*, and *tRNA-Glu* was found so far only in one representative of Procellariiformes (Gibb et al. 2013), Passeriformes (Shi et al. 2017), and Psittaciformes (Urantowska et al. 2018). Unfortunately, selection forces that lead to the more common rearrangement are not clear. It is also possible that a mechanism of duplication and structural organization of the surroundings of control region causes this convergence.

Concerted Evolution of Mitogenomic Duplications in Accipitriformes

Although the reconstruction of ancestral states indicated that the common ancestor of Cathartiformes and Accipitriformes contained a duplication in mitogenomes, the high similarity of duplicated CRs and their phylogenetic analyses could suggest

that the duplication occurred quite recently in five studied species of Cathartiformes as well as early diverged lineages of Accipitriformes. However, such a big number of independent duplication events is highly improbable. Thus, it is more likely that the duplicated regions originated in the common ancestor of Cathartiformes and Accipitriformes or even earlier via tandem duplication followed by random loss (TDRL) (Boore 2000) and constantly persisted in the descendants. However, from time to time, their sequences were homogenized in the way of concerted evolution (fig. 11). This process was proposed for mitochondrial genomes of many other avian species (Kumazawa et al. 1996, 1998; Arndt and Smith 1998; Eberhard et al. 2001; Abbott et al. 2005; Shao et al. 2005; Gibb et al. 2007; Kurabayashi et al. 2008; Cadahia et al. 2009; Eda et al. 2010; Morris-Pocock et al. 2010; Sammler et al. 2011; Schirtzinger et al. 2012; Mackiewicz et al. 2019). We also confirmed the concerted evolution of CRs in owl species.

Interestingly, our phylogenetic analyses including diverged CR2s copies demonstrated that the homogenization of duplicated copies occurred not only in the recently evolving species but also earlier and several times in the evolution of Accipitriformes (fig. 11). A larger number of substitutions

accumulated in CR2 sequences than CR1s suggests that the first copies were subjected to a stronger selection and were used to “overwrite” the degenerated second copies. The most ancient homogenization of CRs occurred after the separation of *Sagittarius serpentarius* and before the divergence of *Pandion haliaetus*. A trace of this event could be represented by CR2 in *Pernis ptilorhynchus*, which was not affected by another homogenization event. Successive sequence uniformization took place after the divergence of *Pernis ptilorhynchus* and before the separation of two main Accipitridae groups, one comprising Circaetinae and Aegypiinae and another including Aquilinae, Haliaeetinae, Milvinae, Buteoninae, Accipitrinae, and Circinae. The lineages of *Aegypius monachus* and *Gyps fulvus* inherited from this event the CR2 that was not subjected to another homogenization in contrast to CR2 in *Spilornis cheela*. Its CRs became similar earlier than in Cathartiformes species as well as *Sagittarius serpentarius* and *Pandion haliaetus*, as indicated by as many as 48% of gaps in their alignment and a greater number of substitutions, which resulted in only 75% identity. In the case of the recent homogenizations, the percent of gaps was much lower, that is, 18–23% and the percent of identity much higher, that is, 92–98%. Two additional homogenizations happened in the common ancestor of Aquilinae and the common ancestor of the rest closely related subfamilies, that is, Haliaeetinae, Milvinae, Buteoninae, Accipitrinae, and Circinae.

Our results are in agreement with those obtained on smaller data set by Cadahia et al. (2009), who also proposed the homogenization of CRs in the common ancestor of Aquilinae. Nevertheless, the concerted evolution could occasionally affect the duplicated fragments in some species of this subfamily (or possible others), because these authors found short and very similar stretches in paralogous copies of CRs in *Aquila heliaca* and *Aquila fasciata*.

A relatively common concerted evolution of control region with adjacent genes can result from a greater vulnerability of this mitogenomic section to duplication and recombination. This can be associated with the role of CR in the initiation of transcription and replication (Boore 1999), which cause that this region stays single-stranded for a longer time. The recombination can be facilitated also by the arrest of replication fork in the region of replication fork barrier (RFB) (Bowmaker et al. 2003; Reyes et al. 2005), in which the 3' end of the nascent strand is exposed for a long time and can be involved in recombination (Kurabayashi et al. 2008).

Besides the mutational point of view, the duplicated fragment can give a selective advantage if an additional control region is involved in the initiation of transcription and replication. The concerted evolution of CRs can cause that their fragments involved in these processes become identical, which can facilitate their common regulation. The processes started from both CR copies can increase the number of genomic and transcript copies per mitochondrion (Kumazawa et

al. 1996; Tang et al. 2000; Umeda et al. 2001). In consequence, it would indicate an increased capacity for energy production by mitochondria. Interestingly, it was found that parrots with two control regions in their mitogenomes are characterized by morphological features associated with more active flight, which consumes a lot of energy (Urantowka et al. 2018). Moreover, it was also noticed that the presence of duplicated control region is connected with increased longevity in various avian groups (Skujina et al. 2016; Urantowka et al. 2018; Mackiewicz et al. 2019).

The selection can be important in Strigiformes, Cathartiformes, Sagittariidae, and Pandionidae, because all of them possess the full second copy of CR. However, the selection likely does not operate on the duplicated fragments in Accipitridae mitogenomes due to the high degeneration of their CR2. It is not inconceivable that in Accipitridae, there is a tendency for elimination of the duplicated regions due to prolongation of the replication time and increasing the energy for the synthesis of the longer genome. The presence of additional copies of duplicated genes, that is, *tRNA-Thr*, *tRNA-Pro*, *ND6*, and *tRNA-Glu*, can be also beneficial for birds possessing them, that is, Cathartiformes, Sagittariidae, and some Strigiformes, but it is disputable because only one functional copy is maintained in other owl species and the members of Accipitriformes. The tandemly duplicated regions were also found in other metazoan mitogenomes (Rand 1993; Wilkinson et al. 1997; Lunt et al. 1998; Xu and Fang 2006; Sun et al. 2015) but their function and association with capacity for energy production can be different. Direct biochemical and physiological studies are necessary to show this relationship.

Many results indicate that the CR homogenization can occur quite often, that is, within individual species lineages at a frequency higher than the rate of speciation. Very similar or identical copies of CRs were reported in species of various bird groups: Bucerotiformes (Sammler et al. 2011), Charadriiformes (Verkuil et al. 2010), Gruiformes (Akiyama et al. 2017), Passeriformes (Gibb et al. 2015; Caparroz et al. 2018; Mackiewicz et al. 2019), Pelecaniformes (Cho et al. 2009; Zhou et al. 2014), Procellariiformes (Abbott et al. 2005; Eda et al. 2010), Psittaciformes (Eberhard et al. 2001; Eberhard and Wright 2016; Urantowka et al. 2018), Strigiformes (Kang et al. 2018), and Suliformes (Morris-Pocock et al. 2010; Gibb et al. 2013). This process can also happen at the individual level in every generation, as found in Bucerotidae (Sammler et al. 2011) and *Kryptolebias marmoratus* (Tatarenkov and Avise 2007). The averaged time between the gene conversions can be approximately 35,000 years as estimated for parrot *Amazona* (Eberhard et al. 2001) or only 850 years in the case of the killifish (Tatarenkov and Avise 2007). We also found cases of the recent homogenization of CRs in the species lineage of Accipitriformes, that is, *Pandion haliaetus* and *Sagittarius serpentarius*, as well as Strigiformes, that is, *Bubo bubo* and *Strix*

uralensis. The large parts of CRs evolved in concert within each lineage of five species of Cathartiformes, that is, *Cathartes aura*, *Cathartes burrovianus*, *Coragyps atratus*, *Sarcoramphus papa*, and *Vultur gryphus*.

Interestingly, there are also cases in which CRs were not homogenized and evolved independently for a quite long time. For example, the different CR copies were maintained longer than the evolution of new species, genera, or families in Falconiformes and various groups of Passeriformes (Mackiewicz et al. 2019). In this study, we also found that the homogenization of CRs did not occur in the present genera of Aquilinae, five Accipitriformes subfamilies (Accipitrinae, Circinae, Buteoninae, Haliaeetinae, and Milvinae), two genera of Aegypiinae as well as *Pernis ptilorhynchus*. Based on the results of molecular dating by Knapp et al. (2019), we can assume that CRs evolved independently in the above-mentioned cases for at least 16.8, 26.7, 33.3, and 46.6 Myr, respectively. The low frequency of CR homogenization in these cases may result from that the D-loop region does not constitute a strong barrier for the replication fork, which decreases the probability of the nascent strand exposition and in consequence recombination of the mitogenome strands (Bowmaker et al. 2003; Reyes et al. 2005; Kurabayashi et al. 2008). It is not inconceivable that an insertion, deletion, or other structural rearrangements made homogenization impossible or rare due to steric hindrance.

Different Evolutionary Pattern of CR Fragments Gives Insight into the Mechanism of Sequence Homogenization

Analyses of duplicated control regions in Cathartiformes revealed that their different parts evolved in various manners, some were subjected to homogenization in a given species lineage, whereas the remaining part did not evolve in concert. The unequal pattern of evolution was observed in CRs of other avian groups, in which a large middle part of CRs and the end of the beginning were homogenized, whereas the beginning of the 5' end and the 3' end of CRs were more variable (Abbott et al. 2005; Cadahia et al. 2009; Cho et al. 2009; Morris-Pocock et al. 2010; Verkuil et al. 2010; Wang et al. 2015). However, the most similar evolutionary pattern observed in *Cathartes* CRs can be found in *Phoebastria albatrosses* (Eda et al. 2010) and representatives of Bucerotidae (Sammler et al. 2011), in which a part evolving in concert is located not only in the middle of CRs but also at their beginning. Thereby, these parts are separated by a more variable fragment that does not show signs of homogenization. This part can correspond to the RFB region, which can halt replication (Reyes et al. 2005; Sammler et al. 2011). According to the model by Kurabayashi et al. (2008), during the stopping at this region, the 3' end of the nascent strand can remain free until its replication restarts and can be easily involved in recombination between two mitogenomic molecules or within a single molecule with a parallel replication fork. The strand

exchange results in the formation of Holliday Junction, which can move toward upstream direction from the 3' ends of nascent strands and stop at the region of poorly matched sequences. Finally, repairing of resultant heteroduplex DNA or secondary replication could lead to the homogenization of recombining parts, including the control region and adjacent genes. It is possible that the second RFB region lies at the variable section of 3' end, where additional recombination can occur.

The presence of the homogenized sections on both sides of the potential RFBs may be associated with bidirectional replication of the mitogenome (Sammler et al. 2011), which was reported for birds (Reyes et al. 2005) and mammals (Bowmaker et al. 2003; Yasukawa et al. 2005; Holt and Reyes 2012; Yasukawa and Kang 2018). According to this replication mode, this process initiates from a broad zone of several kilobases around CR, where a termination region can be also located. In the case of mammalian mitogenome, O_H region appears to function as a RFB and that syntheses of the leading and lagging strands are synchronous (coupled) (Yasukawa and Kang 2018).

Materials and Methods

Samples and DNA Extraction

Blood samples from *Sagittarius serpentarius* and five Cathartiformes species were collected from European zoological gardens (table 1). Total DNA was extracted with Sherlock AX Kit (A&A Biotechnology) according to the manufacturer's protocol.

PCR Strategy for Tandem Duplication Survey

To verify the presence or the absence of tandem duplication within the mitochondrial genomes of *Sagittarius serpentarius* and Cathartiformes representatives, we used the strategy that was proposed by Gibb et al. (2007) and successfully applied for parrot mitogenomes (Urantowska et al. 2018). We designed appropriate D-F and D-R primers to amplify potential fragments between two control regions (fig. 3). Their mutual location within the central part of control regions makes the PCR strategy diagnostic, because the expected amplicons occur only when two control regions are present in the genome. Due to the high variability of the control region sequences between *Sagittarius serpentarius* and *Cathartes aura*, it was impossible to use universal primers. It forced us to design tax-specific primers based on the reference mitogenomes deposited in GenBank (KF961184.1, AY463690.1). Because mitogenomes of other Cathartiformes have not yet been sequenced, primers designed for *Cathartes aura* were used to test the presence of tandem duplication within the mitogenomes of *Cathartes burrovianus*, *Coragyps atratus*, *Sarcoramphus papa*, and *Vultur gryphus*. Based on the selected primers, we ran nine different reactions for each of

the analyzed species (supplementary table S7, Supplementary Material online).

PCR Strategy for Amplification of the Mitogenomic Fragments Containing the Whole Duplicated Regions

The diagnostic fragment designed for tandem duplication survey comprises the second part of the first control region (CR1) and the first part of the second control region (CR2), as well as genes located between them. Therefore, appropriate PCR reactions were performed (fragments 5 and 6 in supplementary table S1, Supplementary Material online) to complete the missing parts of the control regions (CRs) and to reveal the order of genes preceding the CR1. The obtained partial CR1 sequences were used to design species-specific CR-R primers (fig. 3) for the amplification of *ND5*/CR1 fragment (in *Cathartes aura*) or *cytb*/CR1 fragments in other tested species. Similarly, the partial CR2 sequences were used to design species-specific CR-F primers (fig. 3) for the amplification of CR2/12S fragments in all examined species. Appropriate ND5-F, *cytb*-F, and 12S-R primers (supplementary table S1, Supplementary Material online) were designed based on the reference mitogenomic sequences of the analyzed taxa, which were deposited in GenBank (AY463690.1, KF961184.1, KX534422.1, KX534417.1, AY426746.1, KX534425.1, AF173575.1, AF494341.1). Suitable elongation times were applied to avoid the amplification of *ND5*/CR2, *cytb*/CR2, and CR1/12S fragments, which would contain two copies of some genes and/or control regions. In the case of all examined taxa, the whole duplicated mitogenomic regions (*ND5*/12S or *cytb*/12S) were amplified in three overlapping fragments (fig. 3). Amplicons 4 and 6 (supplementary table S1, Supplementary Material online), which contained only one fragment of control region (CR1 or CR2) were 1.6–4.3 kb in length, which excludes the possibility of NUMTs amplification, whose average size is usually below 1 kb (Richly and Leister 2004). Similarly, the length of diagnostic amplicons (fragment 5 in supplementary table S1, Supplementary Material online) located between two control regions was also longer than 1 kb, that is, 2.3–2.7 kb.

PCR Strategy for Amplification of the Remaining Fragments of Mitogenomes

The remaining region of *Cathartes aura* and *Sagittarius serpentarius* mitogenomes, that is, the fragment from 12S gene to *cytb* gene, was not amplified in our study. It was caused by the fact that the mitogenomes of both species were already previously sequenced (Slack et al. 2007; Mahmood et al. 2014) and deposited in GenBank (AY463690.1; KF961184.1). Moreover, the obtained fragments *ND5*/CR, *cytb*/CR, and CR/12S for *Cathartes aura* and *Sagittarius serpentarius* occurred nearly identical with the previously published sequences.

Cathartes aura mitogenome is so far the only complete sequence available for Cathartiformes. Therefore, to design primers useful for the amplification of missing 12S/*cytb* regions, we had to use *cytb*/12S fragments obtained for *Cathartes burrovianus*, *Coragyps atratus*, *Sarcoramphus papa*, and *Vultur gryphus* as references. The set of selected primers (supplementary table S2, Supplementary Material online) enabled us to run 16 different reactions to obtain the missing region as one long amplicon (fragments L in supplementary table S1, Supplementary Material online). Finally, 12S/*cytb* sequences obtained for *Coragyps atratus* and *Sarcoramphus papa* were used as references to design new sets of primers, which enabled us to run reactions appropriate to amplify 12S/*cytb* regions from *Cathartes burrovianus* and *Vultur gryphus* in three overlapping sequences (fragments 1–3 in supplementary table S1, Supplementary Material online).

DNA Amplification and Sequencing

All PCR amplifications were performed in a 25 μ l reaction mixture containing 50 ng of the DNA template. Fragments L (supplementary table S2, Supplementary Material online) from *Sarcoramphus papa* and *Coragyps atratus* were amplified in reaction mixtures, which additionally contained 1U LA Taq Polymerase (TaKaRa; RR002M), 2.5 μ l of 10 \times buffer, 2.5 μ l of MgCl₂, 4 μ l of 10 mM dNTPs, and 0.2 μ l of each primer (10 μ M). All other fragments (1–6 in supplementary table S2, Supplementary Material online) were amplified with addition of 1U DreamTaq Green DNA Polymerase (Thermo Fisher Scientific), 2.5 μ l of 10 \times buffer, 0.6 μ l of 10 mM dNTPs, and 0.6 μ l of each primer (10 μ M). In the case of fragments L, the reaction conditions were as follows: 94 $^{\circ}$ C for 1 min; 98 $^{\circ}$ C for 30 s and 68 $^{\circ}$ C for 15 min repeated 30 times; and 72 $^{\circ}$ C for 5 min. Fragments 1–4 were amplified with the use of the following program: 94 $^{\circ}$ C for 5 min; 94 $^{\circ}$ C for 30 s, 58 $^{\circ}$ C for 30 s, 72 $^{\circ}$ C for 180 s repeated 35 times; and 72 $^{\circ}$ C for 5 min. For the diagnostic fragments 5 (CR1/CR2) as well as fragments 6, the elongation time was shortened and the reaction conditions were as follows: 94 $^{\circ}$ C for 5 min; 94 $^{\circ}$ C for 30 s, 58 $^{\circ}$ C for 30 s, 72 $^{\circ}$ C for 90 s repeated 35 times; and 72 $^{\circ}$ C for 5 min.

For each fragment, the appropriate amount of the PCR reaction mixtures was cleaned with the use of Clean-up Kit (A&A Biotechnology) to obtain the final volume of 100 μ l with the concentration of at least 50 ng/ μ l. Two DNA strands of the cleaned PCR products were sequenced using Primer Walking method (Wyzer Biosciences Inc., Cambridge, MA). Sanger sequencing was carried out on ABI 3730xl machines.

Overlaps between fragments amplified for each species were sufficient to assemble the complete mitogenome of *Cathartes burrovianus*, *Coragyps atratus*, *Sarcoramphus papa*, and *Vultur gryphus*, or whole mitogenomic regions containing duplicated elements in *Cathartes aura* and *Sagittarius serpentarius* with the use of *Ketupa blackistoni*

and *Ketupa flavipes* reference mitogenomes (LC099104, LC099100.1). These two Strigiformes species were chosen because their mitogenomes are characterized by the presence of GO-FD gene order (fig. 1 and [supplementary fig. S5, Supplementary Material](#) online) and they are also quite closely related to Cathartiformes and Accipitriformes taxa (Houde et al. 2019). The annotation of genes was performed in MITOS (Bernt et al. 2013).

Phylogenetic Analyses of Accipitriformes Based on Complete Mitogenomes

To reveal phylogenetic relationships within Accipitriformes, we analyzed 45 complete mitochondrial genome sequences ([supplementary table S8, Supplementary Material](#) online). This set included 33 representatives of Accipitriformes, five members of Cathartiformes, and three species of Strigiformes, which were used as an outgroup. The analyses were based on 13 protein-coding genes (PCGs) as well as two rRNAs and 19 tRNAs genes (not subjected to duplication), which were aligned with Muscle algorithm (Edgar 2004) and concatenated into one alignment with the length of 15,131 bp. We considered 42 partitions of the alignment: three codon positions for each PCG, as well as genes for 12S rRNA, 16S rRNA, and tRNAs.

We applied four methods: Bayesian Approach (BA) implemented in MrBayes (Ronquist et al. 2012) and PhyloBayes (Lartillot et al. 2013) as well as Maximum Likelihood (ML) algorithm implemented in IQ-TREE (Nguyen et al. 2015) and morePhyML (Guindon et al. 2010; Criscuolo 2011). We used PartitionFinder (Lanfear et al. 2017) to evaluate the proper partition scheme for the analyzed alignment. According to AICc criterion, the software proposed 29 different subsets as the best partition ([supplementary table S9, Supplementary Material](#) online), which were used in MrBayes, and IQ-TREE.

In MrBayes, we ran two independent runs with eight chains, each sampled every 100th generation for 30 million generations. Burn-in excluded the first 25% of trees. Remaining trees were used to build a posterior consensus tree after getting the convergence, that is, the standard deviation of split frequencies was below 0.01 after the first 3 million generations and finally reached 0.003 at the end of analysis. In PhyloBayes, we used CAT-GTR+ Γ 5 model with two independent chain runs carried out for 150,000 steps. The first 50,000 generations in each chain were cut as a burn-in and the last 100,000 trees were used to create a consensus tree. Maxdiff value for this analysis was below 0.012. In IQ-TREE, we used nonparametric bootstrap analysis with 1,000 replicates and the increased set of candidate trees equal to 15. In morePhyML, we used GTR+ Γ 4+I+F model along with Shimodara–Hasegawa-like approximate likelihood ratio test (SH-aLRT) (Anisimova and Gascuel 2006) for the estimation of branch support.

Mapping of Mitochondrial Gene Orders on the Phylogenetic Tree

The presence of different mitochondrial gene orders (GO-FD, GO-II, GO-IV) found in Accipitriformes were mapped on the MrBayes tree using Mesquite (Maddison and Maddison 2017) with maximum likelihood (ML) and maximum parsimony (MP) reconstruction methods. AIC criterion was used to select better fit model used in ML method, that is, Mk1 (Markov k-state 1 parameter model).

Clustering Sequences of Control Regions

Forty-five pairs of sequences of the first and the second control regions from 38 representatives of Accipitriformes, five members of Cathartiformes, and two species of Strigiformes were subjected to clustering in CLANS (Cluster Analysis of Sequences) (Frickey and Lupas 2004). This software visualizes BLAST pairwise sequence similarities in either 2D or 3D space. Analyzed sequences are represented in the graph by vertices which are connected by edges reflecting attractive forces proportional to the negative logarithm of the HSP's (high scoring segment pairs) P value. In the pairwise BLAST searches, we used the minimum value of word size equal to 4 and filter for low-complexity region of query sequences.

Preparing Data Sets of Control Regions

Phylogenetic relationships between control region sequences were studied in various data sets. To avoid the artificial grouping of highly diverged and unrelated sequences of the second control regions (CR2s), that is, the long-branch attraction (LBA) artifact (Felsenstein 1978), the individual sequences or those from closely related taxa were compared separately in phylogenetic analyses with the first control regions (CR1s) and other CR2s that were almost identical to their CR1s and did not show the high divergence rate. Moreover, we used DustMasker (Morgulis et al. 2006) to identify and mask out low complexity and repeated sequences in CRs to eliminate clustering of the sequences due to nonspecific regions. Taking into account this procedure, we considered three types of alignments, including: all sites (unmasked), masked sites after aligning the sequences and deprived of the masked sites before the aligning.

The sequences were aligned using M-Coffee, a meta-method assembling multiple sequence alignments obtained from eight individual methods into one single alignment (Wallace et al. 2006). Poorly aligned regions were removed using trimAl applying the best automated method (Capella-Gutierrez et al. 2009). Very diverged and short sequences were removed from the alignments and the whole procedure was repeated again until all such sequences were eliminated. The inclusion of all these combinations resulted in 76 alignments, which were subjected to phylogenetic studies ([supplementary table S6, Supplementary Material](#) online). The

alignments included 40–66 sequences and had the length of 355–1,536 bp. Similarly, we analyzed separately phylogenetic relationships only between ten CRs from Cathartiformes based on three alignment types: including all sites (1,028 bp) as well as with repeats masked (1,109 bp) or removed before aligning the sequences (1,039 bp). All alignments were inspected in JalView (Waterhouse et al. 2009).

Phylogenetic Analyses of Control Regions

We applied four approaches in the inferring of phylogenetic relationships between control regions: two Bayesian analyses in MrBayes (Ronquist et al. 2012) and PhyloBayes (Lartillot et al. 2009) as well as the maximal likelihood method in IQ-TREE (Nguyen et al. 2015) and morePhyML (Guindon et al. 2010; Criscuolo 2011). In morePhyML analyses, we used the best fitted substitution models proposed according to jModelTest (Darriba et al. 2012), whereas in IQ-TREE based on the associated ModelFinder program (Kalyaanamoorthy et al. 2017) (supplementary table S6, Supplementary Material online). In MrBayes, we applied the covarion model assuming various rates of substitutions on different branches (Tuffley and Steel 1998) and mixed models rather than fixed ones to specify appropriate substitution models across the large parameter space (Huelsenbeck et al. 2004) as well as the gamma-distributed rate variation describing heterogeneity rate across sites with five categories as proposed by jModelTest. In PhyloBayes, we applied also the covarion model with the GTR+ Γ 5 model.

Phylogenetic trees were calculated in IQ-TREE using a thorough and slower nearest neighbor interchange (NNI) tree search considering all possible NNIs. In morePhyML, we applied the best heuristic search algorithm, NNI, and subtree pruning and regrafting (SPR). To assess the significance of tree nodes, we used Shimodaira–Hasegawa-like approximate likelihood ratio test (SH-aLRT) with 10,000 replicates and non-parametric bootstrap with 1,000 replicates in these two programs.

In MrBayes, we applied two independent runs starting from random trees, each using 8 and 64 (depending on the alignment set) Markov chains. The trees were sampled every 100 generations for 10 million generations. In the final analysis, we selected trees from the last 2,099,000–7,707,000 (depending on the alignment set) generations that reached the stationary phase and convergence, that is, when the standard deviation of split frequencies stabilized and was much below the recommended threshold 0.01. In PhyloBayes, two independent Markov chains were run for 100,000 steps with one tree sampled for each generation. The last 10,000–90,000 trees (depending on the alignment set) from each chain were collected to compute posterior consensus trees after obtaining convergence, when the largest discrepancy observed across all bipartitions (maxdiff) was much below the proposed threshold 0.1.

Using IQ-TREE, we calculated the consensus of trees obtained in four approaches. The number of the trees that produced a given node were presented together with support values. Tests of tree topologies assuming alternative location of considered CR2s were conducted in IQ-TREE assuming 1 million replicates. We included the following approaches in the testing: RELL method, Shimodaira–Hasegawa test, weighted Shimodaira–Hasegawa test, Expected Likelihood Weight test, and an approximately unbiased test. The topologies that obtained values smaller than 0.05 were considered significantly worse than the best ones. Log-likelihood values for individual 1,496 alignment sites of Cathartiformes control regions were calculated in IQ-TREE under the best fitted model K3Pu+F+G4 for various tree topologies showing different relationships between CRs of Cathartiformes. The length of tree branches leading to CRs copies in individual species was calculated in TempEst (Rambaut et al. 2016).

The program needle from EMBOSS package (Rice et al. 2000) was used to calculate the optimal global alignment of two paralogous sequences based on the Needleman–Wunsch alignment algorithm. The percent of identity was calculated as the percent of identical nucleotides per the number of all nucleotides in the alignment. Additionally, the percent of gaps in the alignment was calculated.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We are also grateful to anonymous reviewers for their insightful comments and remarks, which increased the quality of the manuscript. This work was supported by the National Science Centre Poland (Narodowe Centrum Nauki, Polska) (Grant No. 2017/25/N/NZ8/01197). Some computations were carried out at the Wrocław Centre for Networking and Supercomputing (Grant No. 307). We are grateful to: Vaidl Antonin (Zoological Garden in Prague); Martin Kaiser and Andreas Pauly (Tierpark Berlin-Friedrichsfelde GmbH); Michał Targowski and Michał Krause (Zoological Garden in Gdańsk); Jaroslav Haimy Hyjánek and Michal Podhrazky (Zoological Garden in Důr Králové); Peš Tomáš and Pokorný Jan (Zoological Garden in Plzeň); Marcin Chrapowicki (Zoological Garden in Warsaw); and Marion POUPÉ-LAURENT (Zoo de la Boissière du Doré) for providing samples of examined species. We are also grateful to Mr Jan Kieleczawa (Wyzer Biosciences Inc., Cambridge, MA) for his patience and commitment in sequencing PCR products rich in G and/or C repeats.

Data Availability

The data underlying this article are available in the article and in its [Supplementary Material](#) online as well as in the GenBank Nucleotide Database and can be accessed with AY463690.1, KF961184.1, KX534422.1, KX534417.1, AY426746.1, KX534425.1, AF173575.1, and AF494341.1.

Literature Cited

- Abbott CL, Double MC, Trueman JW, Robinson A, Cockburn A. 2005. An unusual source of apparent mitochondrial heteroplasmy: duplicate mitochondrial control regions in *Thalassarche* albatrosses. *Mol Ecol*. 14(11):3605–3613.
- Akiyama T, et al. 2017. Gene duplication and concerted evolution of mitochondrial DNA in crane species. *Mol Phylogenet Evol*. 106:158–163.
- Anisimova M, Gascuel O. 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol*. 55(4):539–552.
- Arndt A, Smith MJ. 1998. Mitochondrial gene rearrangement in the sea cucumber genus *Cucumaria*. *Mol Biol Evol*. 15(8):1009–1016.
- Bernt M, Braband A, Schierwater B, Stadler PF. 2013. Genetic aspects of mitochondrial genome evolution. *Mol Phylogenet Evol*. 69(2):328–338.
- Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Res*. 27(8):1767–1780.
- Boore JL. 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. In: Sankoff D, Nadeau JH, editors. *Comparative genomes: empirical and analytical approaches to gene order dynamics, map alignment and the evolution of gene families*. Dordrecht (The Netherlands): Kluwer Academic Publishers. p. 133–148.
- Boore JL. 2006. The use of genome-level characters for phylogenetic reconstruction. *Trends Ecol Evol*. 21(8):439–446.
- Boore JL, Brown WM. 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Curr Opin Genet Dev*. 8(6):668–674.
- Boore JL, Fuerstenberg SI. 2008. Beyond linear sequence comparisons: the use of genome-level characters for phylogenetic reconstruction. *Philos Trans R Soc Lond B Biol Sci*. 363(1496):1445–1451.
- Bowmaker M, et al. 2003. Mammalian mitochondrial DNA replicates bidirectionally from an initiation zone. *J Biol Chem*. 278(51):50961–50969.
- Cadahia L, et al. 2009. Repeated sequence homogenization between the control and pseudo-control regions in the mitochondrial genomes of the subfamily Aquilinae. *J Exp Zool B Mol Dev Evol*. 312B(3):171–185.
- Caparroz R, et al. 2018. Mitogenomes of two neotropical bird species and the multiple independent origin of mitochondrial gene orders in Passeriformes. *Mol Biol Rep*. 45(3):279–285.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.
- Cho HJ, et al. 2009. Tandem duplication of mitochondrial DNA in the black-faced spoonbill, *Platalea minor*. *Genes Genet Syst*. 84(4):297–305.
- Clements JF, et al. 2019. The eBird/Clements checklist of birds of the world: v2019. Cornell University, Cornell Lab of Ornithology. Available from: <http://www.birds.cornell.edu/clementschecklist/download/>
- Cooke GM, King AG, Johnson RN, Boles WE, Major RE. 2012. Rapid characterization of mitochondrial genome rearrangements in Australian songbirds using next-generation sequencing technology. *J Hered*. 103(6):882–886.
- Crisuolo A. 2011. morePhyML: improving the phylogenetic tree space exploration with PhyML 3. *Mol Phylogenet Evol*. 61(3):944–948.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 9(8):772.
- Desjardins P, Morais R. 1990. Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. *J Mol Biol*. 212(4):599–634.
- Eberhard JR, Wright TF. 2016. Rearrangement and evolution of mitochondrial genomes in parrots. *Mol Phylogenet Evol*. 94(Pt A):34–46.
- Eberhard JR, Wright TF, Bermingham E. 2001. Duplication and concerted evolution of the mitochondrial control region in the parrot genus *Amazona*. *Mol Biol Evol*. 18(7):1330–1342.
- Eda M, Kuro-o M, Higuchi H, Hasegawa H, Koike H. 2010. Mosaic gene conversion after a tandem duplication of mtDNA sequence in Diomedidae (Albatrosses). *Genes Genet Syst*. 85(2):129–139.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32(5):1792–1797.
- Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool*. 27(4):401–410.
- Formenti G, et al. 2021. Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. *Genome Biol*. 22(1):120.
- Frickey T, Lupas A. 2004. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20(18):3702–3704.
- Gibb GC, et al. 2015. New Zealand Passerines help clarify the diversification of major songbird lineages during the oligocene. *Genome Biol Evol*. 7(11):2983–2995.
- Gibb GC, Kardailsky O, Kimball RT, Braun EL, Penny D. 2007. Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. *Mol Biol Evol*. 24(1):269–280.
- Gibb GC, Kennedy M, Penny D. 2013. Beyond phylogeny: pelecaniiform and ciconiiform birds, and long-term niche stability. *Mol Phylogenet Evol*. 68(2):229–238.
- Grealy A, Bunce M, Holleley CE. 2019. Avian mitochondrial genomes retrieved from museum eggshell. *Mol Ecol Resour*. 19(4):1052–1062.
- Guan XJ, Xu J, Smith EJ. 2016. The complete mitochondrial genome sequence of the budgerigar, *Melopsittacus undulatus*. *Mitochondrial DNA A DNA Mapp Seq Anal*. 27(1):401–402.
- Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 59(3):307–321.
- Hanna ZR, et al. 2017. Complete mitochondrial genome sequences of the northern spotted owl (*Strix occidentalis caurina*) and the barred owl (*Strix varia*; Aves: Strigiformes: Strigidae) confirm the presence of a duplicated control region. *PeerJ* 5:e3901.
- He XL, Ding CQ, Han JL. 2013. Lack of structural variation but extensive length polymorphisms and heteroplasmic length variations in the mitochondrial DNA control region of highly inbred Crested Ibis, *Nipponia nippon*. *PLoS One* 8(6):e66324.
- Holt IJ, Reyes A. 2012. Human mitochondrial DNA replication. *Cold Spring Harb Perspect Biol*. 4(12):a012971.
- Houde P, Braun EL, Narula N, Minjares U, Mirarab S. 2019. Phylogenetic signal of indels and the neoavian radiation. *Diversity* 11(7):108.
- Huang ZH, Tu FY, Ke DH. 2016. Complete mitochondrial genome of blue-throated bee-eater *Merops viridis* (Coraciiformes: Meropidae) with its taxonomic consideration. *Pak J Zool*. 49(1):79–84.
- Huelsenbeck JP, Larget B, Alfaro ME. 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol Biol Evol*. 21(6):1123–1133.
- Jarvis ED, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346(6215):1320–1331.

- Jiang L, et al. 2015. The mitochondrial genomes of *Aquila fasciata* and *Buteo lagopus* (Aves, Accipitriformes): sequence, structure and phylogenetic analyses (vol 10, e0136297, 2015). *PLoS One* 10(10):e0141037.
- Jiang L, et al. 2019. Complete mitochondrial genome sequence of the Himalayan Griffon, *Gyps himalayensis* (Accipitriformes: Accipitridae): sequence, structure, and phylogenetic analyses. *Ecol Evol.* 9(15):8813–8828.
- Johnson JA, Brown JW, Fuchs J, Mindell DP. 2016. Multi-locus phylogenetic inference among New World Vultures (Aves: Cathartidae). *Mol Phylogenet Evol.* 105:193–199.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 14(6):587–589.
- Kang H, Li B, Ma X, Xu Y. 2018. Evolutionary progression of mitochondrial gene rearrangements and phylogenetic relationships in Strigidae (Strigiformes). *Gene* 674:8–14.
- Knapp M, et al. 2019. Mitogenomic evidence of close relationships between New Zealand's extinct giant raptors and small-sized Australian sister-taxa. *Mol Phylogenet Evol.* 134:122–128.
- Kumazawa Y, Ota H, Nishida M, Ozawa T. 1996. Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. *Mol Biol Evol.* 13(9):1242–1254.
- Kumazawa Y, Ota H, Nishida M, Ozawa T. 1998. The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. *Genetics* 150(1):313–329.
- Kurabayashi A, et al. 2008. Phylogeny, recombination, and mechanisms of stepwise mitochondrial genome reorganization in mantellid frogs from Madagascar. *Mol Biol Evol.* 25(5):874–891.
- Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol.* 34(3):772–773.
- Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25(17):2286–2288.
- Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol.* 62(4):611–615.
- Lavrov DV. 2007. Key transitions in animal evolution: a mitochondrial DNA perspective. *Integr Comp Biol.* 47(5):734–743.
- Lavrov DV. 2011. Key transitions in animal evolution: a mitochondrial DNA perspective. In: Desalle R, Schierwater B, editors. Key transitions in animal evolution. Boca Raton (FL): CRC Press. p. 34–53.
- Lerner HRL, Klaver MC, Mindell DP. 2008. Molecular phylogenetics of the Buteonine birds of prey (Accipitridae). *Auk* 125(2):304–315.
- Liu G, Li C, Du Y, Liu X. 2017. The complete mitochondrial genome of Japanese sparrowhawk (*Accipiter gularis*) and the phylogenetic relationships among some predatory birds. *Biochem Syst Ecol.* 70:116–125.
- Lounsbury ZT, et al. 2015. Next-generation sequencing workflow for assembly of nonmodel mitogenomes exemplified with North Pacific albatrosses (*Phoebastria* spp.). *Mol Ecol Resour.* 15(4):893–902.
- Lunt DH, Whipple LE, Hyman BC. 1998. Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. *Mol Ecol.* 7(11):1441–1455.
- Mackiewicz P, Urantowka AD, Krocak A, Mackiewicz D. 2019. Resolving phylogenetic relationships within passeriformes based on mitochondrial genes and inferring the evolution of their mitogenomes in terms of duplications. *Genome Biol Evol.* 11(10):2824–2849.
- Maddison WP, Maddison DR. 2017. Mesquite: a modular system for evolutionary analysis. Version 3.31. Available from: <http://www.mesquiteproject.org>
- Mahmood MT, McLenachan PA, Gibb GC, Penny D. 2014. Phylogenetic position of avian nocturnal and diurnal raptors. *Genome Biol Evol.* 6(2):326–332.
- Mindell DP, Fuchs JJ, Johnson JA. 2018. Phylogeny, taxonomy and geographic diversity of diurnal raptors: Falconiformes, Accipitriformes, and Cathartiformes. In: Sarasola JH, Grande JM, Negro JJ, editors. Birds of prey. Birds of prey biology and conservation in the XXI century. Cham (Switzerland): Springer. p. 3–32.
- Mindell DP, Sorenson MD, Dimcheff DE. 1998. Multiple independent origins of mitochondrial gene order in birds. *Proc Natl Acad Sci U S A.* 95(18):10693–10697.
- Morgulis A, Gertz EM, Schaffer AA, Agarwala R. 2006. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J Comput Biol.* 13(5):1028–1040.
- Morris-Pocock JA, Taylor SA, Birt TP, Friesen VL. 2010. Concerted evolution of duplicated mitochondrial control regions in three related seabird species. *BMC Evol Biol.* 10:14.
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.
- Oatley G, Simmons RE, Fuchs J. 2015. A molecular phylogeny of the harriers (Circus, Accipitridae) indicate the role of long distance dispersal and migration in diversification. *Mol Phylogenet Evol.* 85:150–160.
- Ogoh K, Ohmiya Y. 2007. Concerted evolution of duplicated control regions within an ostracod mitochondrial genome. *Mol Biol Evol.* 24(1):74–78.
- Pratt RC, et al. 2009. Toward resolving deep neaves phylogeny: data, signal enhancement, and priors. *Mol Biol Evol.* 26(2):313–326.
- Prum RO, et al. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526(7574):569–573.
- Qin XM, Guan QX, Shi JP, Hou LX, Qin PS. 2013. Complete mitochondrial genome of the *Spilornis cheela* (Falconiformes, Accipitridae): comparison of *S. cheela* and *Spizaetus alboniger*. *Mitochondrial DNA* 24(3):255–256.
- Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2(1):vew007.
- Rand DM. 1993. Endotherms, ectotherms, and mitochondrial genome-size variation. *J Mol Evol.* 37(3):281–295.
- Reyes A, Yang MY, Bowmaker M, Holt J. 2005. Bidirectional replication initiates at sites throughout the mitochondrial genome of birds. *J Biol Chem.* 280(5):3242–3250.
- Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16(6):276–277.
- Richly E, Leister D. 2004. NUMTs in sequenced eukaryotic genomes. *Mol Biol Evol.* 21(6):1081–1084.
- Rodrigues P, A'lvarez P, Verdugo C. 2017. Complete mitochondrial genome of the Neotropic cormorant (*Phalacrocorax brasilianus*). *Mitochondrial DNA A DNA Mapp Seq Anal.* 28(3):401–402.
- Rokas A, Holland PW. 2000. Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol.* 15(11):454–459.
- Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61(3):539–542.
- Roques S, Godoy JA, Negro JJ, Hiraldo F. 2004. Organization and variation of the mitochondrial control region in two vulture species, *Gypaetus barbatus* and *Neophron percnopterus*. *J Hered.* 95(4):332–337.
- Sammler S, Bleidorn C, Tiedemann R. 2011. Full mitochondrial genome sequences of two endemic Philippine hornbill species (Aves: Bucerotidae) provide evidence for pervasive mitochondrial DNA recombination. *BMC Genomics* 12:35.

- Schirtzinger EE, et al. 2012. Multiple independent origins of mitochondrial control region duplications in the order Psittaciformes. *Mol Phylogenet Evol.* 64(2):342–356.
- Shao R, Barker SC, Mitani H, Aoki Y, Fukunaga M. 2005. Evolution of duplicate control regions in the mitochondrial genomes of metazoa: a case study with Australasian Ixodes ticks. *Mol Biol Evol.* 22(3):620–629.
- Shi RR, Chen K, Li SB. 2017. A novel gene organization of the rock sparrow *Petronia petronia* (Aves: Passeriformes) revealed by complete mitochondrial genome. *Mitochondrial DNA B Resour.* 2(2):858–859.
- Singh TR, Shneor O, Huchon D. 2008. Bird mitochondrial gene order: insight from 3 warbler mitochondrial genomes. *Mol Biol Evol.* 25(3):475–477.
- Skujina I, McMahon R, Lenis VPE, Gkoutos GV, Hegarty M. 2016. Duplication of the mitochondrial control region is associated with increased longevity in birds. *Aging* 8(8):1781–1789.
- Slack KE, Delsuc F, McLenachan PA, Arnason U, Penny D. 2007. Resolving the root of the avian mitogenomic tree by breaking up long branches. *Mol Phylogenet Evol.* 42(1):1–13.
- Song X, et al. 2015. The complete mitochondrial genome of *Accipiter virgatus* and evolutionary history of the pseudo-control regions in Falconiformes. *Biochem Syst Ecol.* 58:75–84.
- Sun S, Kong L, Yu H, Li Q. 2015. The complete mitochondrial DNA of *Tegillarca granosa* and comparative mitogenomic analyses of three Arcidae species. *Gene* 557(1):61–70.
- Tamashiro RA, et al. 2019. What are the roles of taxon sampling and model fit in tests of cyto-nuclear discordance using avian mitogenomic data? *Mol Phylogenet Evol.* 130:132–142.
- Tang Y, Manfredi G, Hirano M, Schon EA. 2000. Maintenance of human rearranged mitochondrial DNAs in long-term cultured transmittochondrial cell lines. *Mol Biol Cell.* 11(7):2349–2358.
- Tatarenkov A, Avise JC. 2007. Rapid concerted evolution in animal mitochondrial DNA. *Proc Biol Sci.* 274(1619):1795–1798.
- Tuffley C, Steel M. 1998. Modeling the covarian hypothesis of nucleotide substitution. *Math Biosci.* 147(1):63–91.
- Umeda S, et al. 2001. Both heavy strand replication origins are active in partially duplicated human mitochondrial DNAs. *Biochem Biophys Res Commun.* 286(4):681–687.
- Urantowska AD, et al. 2018. New insight into parrots' mitogenomes indicates that their ancestor contained a duplicated region. *Mol Biol Evol.* 35(12):2989–3009.
- Urantowska AD, Krocak A, Mackiewicz P. 2020. New view on the organization and evolution of Palaeognathae mitogenomes poses the question on the ancestral gene rearrangement in Aves. *BMC Genomics* 21(1):874.
- Verkuil YI, Piersma T, Baker AJ. 2010. A novel mitochondrial gene order in shorebirds (Scolopacidae, Charadriiformes). *Mol Phylogenet Evol.* 57(1):411–416.
- Wallace IM, O'Sullivan O, Higgins DG, Notredame C. 2006. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res.* 34(6):1692–1699.
- Wang HW, et al. 2016. The whole mitochondrial genome of the Lesser Kestrel (*Falco naumanni*). *Mitochondrial DNA A DNA Mapp Seq Anal.* 27(4):2385–2386.
- Wang XY, Huang Y, Liu N, Yang J, Lei FM. 2015. Seven complete mitochondrial genome sequences of bush tits (Passeriformes, Aegithalidae, Aegithalos): the evolution pattern in duplicated control regions. *Mitochondrial DNA* 26(3):350–356.
- Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189–1191.
- Wilkinson GS, Mayer F, Kerth G, Petri B. 1997. Evolution of repeated sequence arrays in the D-loop region of bat mitochondrial DNA. *Genetics* 146(3):1035–1048.
- Xu QH, Fang SG. 2006. Variable number tandem repeats in the mitochondrial DNA control region of the Chinese alligator, *Alligator sinensis*. *Amphib Reptilia.* 27(1):93–101.
- Yasukawa T, Kang D. 2018. An overview of mammalian mitochondrial DNA replication mechanisms. *J Biochem.* 164(3):183–193.
- Yasukawa T, Yang MY, Jacobs HT, Holt JJ. 2005. A bidirectional origin of replication maps to the major noncoding region of human mitochondrial DNA. *Mol Cell.* 18(6):651–662.
- Zhang LT, Zhang M, He SY. 2017. The complete mitochondrial genome of great cormorant, *Phalacrocorax carbo* (Phalacrocorax, Phalacrocoracidae). *Mitochondrial DNA A DNA Mapp Seq Anal.* 28(1):1–2.
- Zhou X, Lin Q, Fang W, Chen X. 2014. The complete mitochondrial genomes of sixteen ardeid birds revealing the evolutionary process of the gene rearrangements. *BMC Genomics* 15:573.

Associate editor: Liliana Milani